

指导教师： 杨涛

提交时间： 2016/3/21

CVPR2015 Paper

Translation

No: 01

姓名： 陈虹志

学号： 2013302598

班号： 10011306

基于自动模型选择的贝叶斯自适应矩阵分解

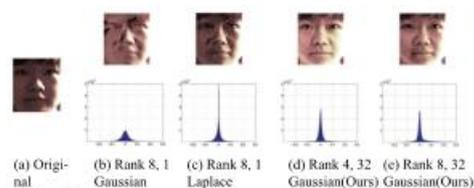
摘要

低秩矩阵分解长期以来被公认为许多计算机视觉应用中的一个基本问题。然而，由于选择的噪声模型和确定模型的能力所带来的挑战，现有的矩阵分解方法的可靠性往往很难保证。我们通过在本文中提出一个非参数贝叶斯自适应矩阵分解 (AMF) 模型解决这些问题。AMF 提出了一个新的基于 Dirichlet 过程混合高斯模型的噪声模型 (dpgmm) 利用其组件的选择和拟合能力广泛的未知噪声高的灵活性。AMF 还规定测定 (ARD) 之前的低等级系数矩阵，秩可确定自动不需要执行任何硬约束。一个高的变分方法，然后设计模型推理。我们比较了 AMF 与国家的最先进的基于数据矩阵分方法设置范围从综合数据的实际应用数据。从结果，AMF 始终达到更好或相当的性能。

1. 简介

矩阵分解是许多的一个重要组成部分计算机视觉应用，如人脸识别

[35]，运动分割 [7]，从运动和结构 (SFM) [27]。简单地说，它近似于一个给定的数据在一些标准的基础矩阵和系数矩阵的乘积矩阵。如果两个潜在的秩系数矩阵是低于原始数据矩阵，矩阵分解是一种有效的方式来揭示数据的低维结构。矩阵分解的成功很大程度上取于在选择适当的模型。两模型选择问题参与。



第一个是噪声模型的选择它如何影响该矩阵的每个条目可以由模型为代表，第二个是选择图 1。一脸的阴影去除实例示范效应一些模型选择问题。原始图像显示 (一) 和不同型号的回收结果显示在 (B) — (E)。在结果中，第一行显示的图像恢复和第二行显示的相应的噪声分布矩阵。对两因子矩阵模型的表达能力相关的能力。在图 1 中，我们的经验表明在脸上的阴影去除这两个问题的影响应用。我们可以看到，确实影响的

噪声模型性能大而等级控制的程度抽象。

对于第一个模型选择问题，各种尝试得到的基本数据的噪声更好的描述，从传统的 ℓ_2 规范的强大的 ℓ_1 规范 [5] 非凸规范 [30]。这些方法往往对噪声过强的假设分布而不幸的是，不持有许多应用。最近，孟和 de la Torre 提出利用高斯混合模型 (GMM) 来拟合噪声 [18]，并进一步扩展到一个完整的贝叶斯模型 [32]。虽然它比上述方法更灵活的在高斯分量数仍事先规定的。这是一个真正的限制自数量的高斯沉重地压在概化一个 GMM 的噪声模型。在本文中，我们提出了使用 Dirichlet 过程混合高斯模型 (dp-gmm) [22] 由于噪声模型。一方面，我们可以利用事实上，GMM 是任何连续分布 [17]，从而能够适应各种类型的万能逼近器噪声。另一方面，我们可以推断出一些从数据所需的高斯组件，而不是做启发式修剪或在中性点不接地的猜测。

对于第二个问题，一种方法是限制级该因子矩阵直接而另一个是征收正则化的人来说。除了一些应用程序如无这强大的真实等级知识是可用的，一般人很难准确估计潜在的秩。一些启发式方法 [26, 19] 有用于排序选择但他们的表现不稳定，在不同的应用程序。对于正则化方法，提出了一些正则化以满足特定应用的要求，如依据正交约束在 SFM 和非

负性在图像分析中的约束。一些人已经选择只是为了避免过拟合，例如，气势 ℓ_2 正则化 [24, 28]，这是核的一个推广规范，有效的减少造成的秩矩阵。在这里我们选择自动关联性测定 (ARD) [15, 20] 之前的系数矩阵。ARD 已长期以来被认为是一种有效的检测技术输入的相关部件，这样我们就可以自动推断出最佳的排名通过修剪其他无关的人。

在本文中，我们提出了我们的新的非参数的完整自适应矩阵分解的贝叶斯模型 (AMF)。AMF 首次充分利用的灵活性和适应能力强的 dp-gmm 噪声模型，并完成通过与自动排序选择。它的设计也能够与缺失数据处理输入。模型推理，我们设计了一个高效的变分方法在 DP 掰棒子表示。实验验证，我们使用文字去除面部的阴影去除任务证明 AMF 的模型自动选择能力的有效性。

2. 相关工作

矩阵分解的数值分析中有它的根。最常用的方法之一是奇异值分解 (SVD)，得到最优解如果真正的排名是已知的加性高斯噪声假设。奇异值分解的一个缺点是它不能处理缺失数据。为了弥补这一缺陷，布坎南和吉本提出阻尼牛顿法 [4]。为了提高 SVD 的鲁棒性，基于迭代加权最小二乘法最小二乘估计是第一个提出 [7]。以下是，其他早期的方法把矩阵分解为几个小问题的线性规划

问题，在各步，例如，[13, 8]。然而，所有这些方法都高对于大规模的计算成本，因此不适合应用。在最近的一次突破，矩阵分解已通过主成分的追求带来 (PCP) [5]。PCP 对一个凸优化问题，问题。它采用核范数规范等级由此产生的矩阵和 ℓ_1 范数作为噪声模型容纳在矩阵值。解决优化问题，主张采用交替方向乘数法 (ADMM) [3] 产生相当大的加速相比，上述方法。受此启发创业，[33, 34, 26] 利用不同一企业设置的矩阵分解，显示出了巨大的在一些应用传统方法的改进。另一个新兴的趋势是采用梯度下降在格拉斯曼流形，例如，[11, 31]。这些方法的优势在于他们的学习能力低秩矩阵分解的在线。在 [24]，salakhutdinov 和 mnih 首先制定了矩阵分解问题在概率框架。然后它被扩展一个完整的贝叶斯模型 [23] 和 [14] 的非线性模型。这三种模式非常适合协同过滤应用。沿着这条线，王等人。提出了两稳健贝叶斯矩阵分解方法从点估计的角度 [28] 和 [29] 全贝叶斯的角度来看，在一些计算机视觉应用中取得了很好的结果。其他一些相关的方法包括贝叶斯稳健主成分分析 [6] [6]，被视为一个概率版的 PCP，和变分贝叶斯低秩矩阵估计 [1]，这也依赖于在 ARD 和使用快速变分推理算法贝叶斯矩阵分解。这是以前的作品最密切相关的我们的 [18, 32]。他们都使用一个 GMM 模型可能是复杂

的和未知的噪声在数据。虽然他们加在后者的工作 [32] 因子矩阵的先验设定数量的需要高斯成分事先在很大程度上限制了该模型的概化。这种新算法缺乏与缺失数据处理输入的能力。我们将这些相关模型的实证比较表明 AMF 始终获得更好的或类似结果。

3. 记号

我们介绍了一些符号被用来在续集。一个矩阵 X ， X^T 表示其转置和 $\text{TR}(x)$ 的微量。我们还用 $i \cdot$ 和 $X \cdot j$ 表示第 i 行和 j 柱，分别为 X 。让我表示单位矩阵适当的尺寸。概率分布， $n(\mu, \Sigma)$ 表示的多元正态分布，均值向量 μ 和协方差矩阵 Σ ， $N(\mu, \sigma)$ 单因素正常平均 μ 和方差 σ 分布， $B(\alpha, \beta)$ 参数 α 和 β 分布，免疫球蛋白 (α, β) 形状参数和尺度参数 $\alpha \beta$ 倒 Gamma 分布，以及多 (π) 的多项式。

4. 背景

在本节中，我们介绍一些背景知识在下一节介绍我们的模型设定阶段。首先介绍了低秩矩阵的一般形式分解然后回顾 Dirichlet 过程。

4.1. 低秩矩阵分解

作为最常见的低秩矩阵分解模型的输入数据矩阵 Y 可以表示为

$$Y = UV^T + E,$$

其中的数据矩阵 $Y = [y_{mn}] \in \mathbb{R}^{M \times n}$ 假定是低等级的， $U \in \mathbb{R}^{M \times R}$ 和 $V \in \mathbb{R}^{N \times R}$ 因子矩阵， $E = [e_{mn}] \in \mathbb{R}^{M \times N}$ 表示添加剂噪声。表示基本上分

解为两个低等级系数矩阵的秩为 r 的 $\ll \min(M, N)$ 。

例如，如果数据只是一个单一的形象，假定是低等级，然后矩阵 Y 以简单被设置为图像。如果数据是一个视频或一组高度相关的图像，那么我们首先需要量化的视频或图像的每一帧集合形成一系列的矩阵 Y 。

4.2. Dirichlet 的过程

因为我们将定义基于 dp-gmm 噪声模型，我们先回顾一下这里 Dirichlet 过程的基本概念和应用。

4.2.1 Dirichlet 过程

Dirichlet 过程 (DP) 的概念，提出了外汇储备—guson [9]。我们可以考虑一个 DP 的或延伸—通过将组件数 k 无限普通 Dirichlet 分布。让我们把 dp-gmm 为例说明。DP 是在一个无限的分布簇数，每一组参数 χ_K 高斯分布的均值和方差，即。从这个 DP 平局将选择一些集 K 根据 Dirichlet 分布的参数并返回特定于群集。因此，我们得到一个与参数设置 χ_K 高斯分布。因此 DP 在分布“分配”。

DP 提供聚类数选择充分的灵活性。任何建立在 DP 模型能够适应多项分布的任意多的类别或可能的结果。在混合模型中，DP 提供了一种非参数贝叶斯选择，能够自动确定模型所需的目标分布混合元素数量。

4.2.2 坚持破除施工

坚持打破建设指的是一个随机

过程构建 DP [25]。考虑一个单位长度的棒

$$\theta_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l),$$

$(0, 1)$ 。我们先画一个价值 β_1 从 Beta 分 $B(1, \alpha)$ 。并挑选一部分其余的。然后我们画 β_2 从 $B(1, \alpha)$ 让 θ_{β_2} 等于 $2(1 - \beta_1)$ 。重复这一过程。我们有一个序列的木棍长度在 β_K 从分发独立绘 $B(1, \alpha)$ 。很容易发现随 $\sum_{k=1}^{\infty} \theta_k = 1$ 。继建设上面的棒断裂分布 θ 写成 θ 创业板 (α) [21]。

5. 我们的模型

在这一部分中，我们首先提出的关键方法动机相关的自适应矩阵分解 (AMF) 模型。这是随后的 AMF 的具体的生成过程。在那之后，我们的关系探讨在 AMF 和现有方法的基础上， ℓ_2 或 l_1 损失。

5.1. 自适应矩阵分解

矩阵分解的成功很大程度上依赖于模型的选择，其中，讨论之前，包括确定的因子矩阵的容量选择噪声模型。

在大多数实际应用中，实际的秩 r 建模所需数据最初是未知的。例如，在一个背景减法任务静态回—理想地，是适当的组的秩 1。然而，如果背景是多模态（例如，由于周期性变化的对象，如交通灯），秩需要的是通常更高。确定合适的等级，常用的方法是尝试不同的值通过多次运行，

然后选择一个 R 得到最佳的性能。为了避免低效毫无根据的尝试，我们采用自动关联性测定 (ARD) 方法 [16] 通过征收之前每个维度 (柱) u 和 v 减少无关列损害性能。具体来说，我们对高斯先验方差 λ R u 和 v 列：

$$\begin{aligned} p(\mathbf{U} | \lambda) &= \prod_{r=1}^R \mathcal{N}(\mathbf{u}_{\cdot r} | 0, \lambda_r \mathbf{I}_M), \\ p(\mathbf{V} | \lambda) &= \prod_{r=1}^R \mathcal{N}(\mathbf{v}_{\cdot r} | 0, \lambda_r \mathbf{I}_N), \end{aligned}$$

在 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_R)$

在建模的噪声 E 方面，虽在使用一个单一的高斯和拉普拉斯分布普遍，这种方法往往是不够的许多现实世界应用中，噪声可以是不同的类型或从不同的来源。因此，简单的使用 ℓ_2 或 ℓ_1 损失不能得到令人满意的结果。考虑到一千二百八十六模型已被证明是一个万能逼近器为任何连续的密度函数，它可以被用来更好地描述噪声来自未知来源的应用可能多的概率分布。例如，当建模的图像，我们可以使用高斯分布的方差大处理大造成的偏差阴影或闭塞而使用另一个高斯分布拟合传感器的噪声方差较小。这在实际的应用提供了更大的灵活性。

事实上，这个想法最近 [18] 追求，[32]。然而，对高斯分量数 k GMM 在 [18, 32] 已被指定了一个先验的。它的价值往往会影响模型的性能显著。如果 K 太小，集群可能模型从广泛的来源复杂的噪声。如果 K 太大，然而，它将采取一切费时高斯分量考虑大多数人贡献不大的噪声建模。此

外，没有一个适当的事先对 GMM 混合系数的模型是不稳定的，容易过度拟合。

针对这些问题，我们使用一个 dp-gmm 这里建模的噪声。一方面，它保留的表达功率 GMM。另一方面，它利用非参数贝叶斯方法使用 DP 从数据中自动确定高斯数。一个有吸引力的利用非参数贝叶斯方法是这样的，而不是强加的假设可能是错误的，它“让数据说话”。

基于 GMM 棒断裂构造，噪声模型走出困境的选择适当数量的组件。放松的 K 后无穷大，噪声 ϵ_{mn} 可以表示为：

$$p(\epsilon_{mn}) = \sum_{k=1}^{\infty} \theta_k \mathcal{N}(\epsilon_{mn} | 0, \sigma_k),$$

在每个高斯成分的混合比例一分从棒断裂的过程，即 $\sum_{k=1}^{\infty} \theta_k = 1$ 。

因此，噪声参赛作品将簇成 k 组无对于复杂的模型选择过程需要。

5.2. 生成过程

我们现在把 dp-gmm ARD 和理想的特征来定义的 AMF 模型。我们也将适当的共轭先验的参数，如适用。图形模型如图 2 所示，AMF 的生成过程如下：

1. 绘制组件的混合比 $\theta \sim \text{GEM}(\alpha)$.
2. 每个噪声聚类：

- 得出方差 $\sigma_k \sim \text{Ig}(0, 0)$

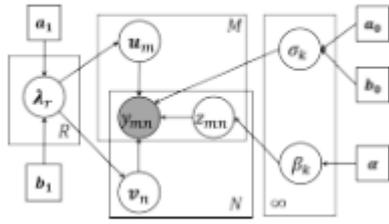


图 2. AMF 的图形模型

3. 对于每个 U 和 V 尺寸（即每列 U 和 V ）：

- 得出方差 $\lambda \sim \text{Ig}(1, 1)$

4. 在 U 和 V 的每个元素：

- 画 $U, V \sim N(0, \lambda)$

5. 为每个数据元素 Y 锰：

- 绘制噪声聚类标签 $Z \sim \text{Dir}(\theta)$

- 画观察 $Y \sim N(\mu_m \cdot V, \sigma^2)$

在这里， $\theta_k \triangleq \beta_k \prod_{l=1}^{k-1} (1 - \beta_l)$ 和 β_K

是相互独立从 $B(1, \alpha)$ 根据棒断裂结构在该模型中，一个 $0, B(0, B(1, 1, \alpha))$ 是超。基在生成过程中，联合分布可以表示为：

$$\begin{aligned} p(\mathbf{U}, \mathbf{V}, \mathbf{Y}, \mathbf{z}, \sigma, \lambda, \beta | a_0, b_0, a_1, b_1, \alpha) \\ = p(\mathbf{Y} | \mathbf{U}, \mathbf{V}, \mathbf{z}, \sigma) p(\mathbf{U} | \lambda) p(\mathbf{V} | \lambda) p(\lambda | a_1, b_1) \\ p(\sigma | a_0, b_0) p(\mathbf{z} | \beta) p(\beta | \alpha). \end{aligned}$$

5.3. 与现有方法的关系

让我们试着探讨所提出的方法是基于 ℓ_2 或 ℓ_1 损失相关的现有的方法。首先，很容易看出，使用 ℓ_2 损失，相当于使用一个单一的高斯拟合噪声，这显然是不灵活的。其次，在 [28] 指出，该 ℓ_1 范数或相应的拉普拉斯

分布可以表示为混合集成的混合高斯分布等对指数分布。这意味着，矩阵中的每个条目，它有自己的 GMM 表示噪声。该方案提供了太多的噪声建模的灵活性因为这是很少的情况下任何噪音影响的只有一个数据录入。我们的方法可以被看作是一个权衡之间这两种方法允许不同的作品分享相同的高斯分布。

6. 变分推断

在贝叶斯参数估计的关键问题模型是计算潜在的后验分布一千二百八十七由于观测到的数据变量。像许多贝叶斯模型，在我们的后验分布的精确推理模型是棘手的，因此近似推理需要。虽然抽样方法如马尔可夫链蒙特卡罗 (MCMC) 算法可以提供非常精确的渐近逼近后，他们往往是缓慢的高维数据。此外，收敛不易察觉。作为一个更高效确定性替代 MCMC 算法，我们采用平均场 AMF 在本文中的变分方法。

变分法近似的潜变量的后验分布的分解形式组成的新的变分分布 Q 为潜变量自由变分参数。逼近了接近目标后验分布尽可能通过最小化 Kullback-Leibler (KL) 发散的两个分布。

基于平均场的变分方法，我们设计了以下变配：

$$\begin{aligned} q(\mathbf{U}, \mathbf{V}, \mathbf{z}, \sigma, \lambda, \beta) = \prod_{m=1}^M q_{u_m}(\mathbf{a}_m, \Sigma_m) \prod_{n=1}^N q_{v_n}(\mathbf{b}_n, \Sigma_n) \\ \prod_{r=1}^R q_{\lambda_r}(\eta_r) \prod_{k=1}^K q_{\beta_k}(\gamma_k) \prod_{k=1}^K q_{\sigma_k}(\tau_k) \prod_{m=1}^M \prod_{n=1}^N q_{z_{mn}}(\phi_{mn}), \end{aligned}$$

在每排 U 服从高斯分布意味着

我 • 和协方差 Σ UM 和它类似的 V, λ 和 σ K 遵循逆 Gamma 分布的参数通过 η R, 1, η R, 2, 和 τ K, 1, 2, 分别 τ K, $Q\beta$ K (γ K) 是一个分布, 和 Q Z 锰 (Mn ϕ) 是一个多项式分布在 [2]的工作, 逼近了截断棒在 K 断施工, 已证明接近真正的 DP 只要 K 选择要足够大, [12]。实证 K 可以初始化从几十到几百的价值的基礎上模型的复杂性。无用的尺寸将逐渐自动修剪。

KL 散度最小化的优化问题是相当于最大以下低绑定:

$$\mathcal{L} = \mathbb{E}_q[\log p(\mathbf{U}, \mathbf{V}, \mathbf{Y}, \mathbf{z}, \sigma, \lambda, \beta)] - \mathbb{E}_q[\log q(\mathbf{U}, \mathbf{V}, \mathbf{z}, \sigma, \lambda, \beta)],$$

其中 $\mathbb{E}_q[\cdot]$ 表示期望就有最大化是通过迭代更新的每个参数, 而保持其他参数固定 L 相对于参数为零的导数进行设置。

6.1. 更新 γ , τ 和 ϕ

我们按照程序设定一个大棍子打破 K 截断近似足够的价值。让 Ω 表示所观察到的数据指标的设置。参数更新如下:

$$\begin{aligned} \gamma_{k,1} &= 1 + \sum_{(m,n) \in \Omega} \phi_{mnk}, \\ \gamma_{k,2} &= \alpha + \sum_{(m,n) \in \Omega} \sum_{t=k+1}^K \phi_{mnt}; \\ \tau_{k,1} &= a_0 + \frac{1}{2} \sum_{(m,n) \in \Omega} \phi_{mnk}, \\ \tau_{k,2} &= b_0 + \frac{1}{2} \sum_{(m,n) \in \Omega} \phi_{mnk} \mathbb{E}_q[(y_{mn} - \mathbf{u}_m \cdot \mathbf{v}_n \cdot T)^2] \end{aligned} \quad (8)$$

for $k = 1, \dots, K$ and $(m, n) \in \Omega$, where

$$\begin{aligned} \phi_{mnk} &\propto \exp\left\{\mathbb{E}_q[\log \beta_k] + \sum_{t=1}^{k-1} \mathbb{E}_q[\log(1 - \beta_t)]\right. \\ &\quad \left. - \frac{1}{2} \frac{\tau_{k,1}}{\tau_{k,2}} \mathbb{E}_q[(y_{mn} - \mathbf{u}_m \cdot \mathbf{v}_n \cdot T)^2]\right. \\ &\quad \left. - \frac{1}{2} [\log \tau_{k,2} - \psi(\tau_{k,1})]\right\}, \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbb{E}_q[(y_{mn} - \mathbf{u}_m \cdot \mathbf{v}_n \cdot T)^2] &= y_{mn}(y_{mn} - 2\mathbf{a}_m \cdot \mathbf{b}_n^T) \\ &\quad + \text{tr}\left((\mathbf{a}_m^T \mathbf{a}_m + \Sigma_m^u)(\mathbf{b}_n^T \mathbf{b}_n + \Sigma_n^v)\right), \\ \mathbb{E}_q[\log \beta_k] &= \psi(\gamma_{k,1}) - \psi(\gamma_{k,1} + \gamma_{k,2}), \\ \mathbb{E}_q[\log(1 - \beta_k)] &= \psi(\gamma_{k,2}) - \psi(\gamma_{k,1} + \gamma_{k,2}). \end{aligned} \quad (10)$$

这里 ψ 表示双函数

备注: 请注意, τ 更新之后, 我们可以很容易的获混合比例 θ 的 k 个簇。归一化后, 我们会修剪这些簇的概率较比一个预定义的阈值, 因为一个很小的 θ K 表对于某些条目属于相应的集群是不太可能的。因此, 不构成打任何重要的角色可以被删除。

6.2. 更新

\mathbf{a}_m , \mathbf{b}_n , Σ_m^u , Σ_n^v and η :

我们的下一个更新到 U 和 V 的相关参数:

$$\begin{aligned} \mathbf{a}_m \cdot T &= \Sigma_m^u \cdot \left(\sum_{n:(m,n) \in \Omega} \sum_{k=1}^K \frac{\tau_{k,1}}{\tau_{k,2}} (y_{mn} \phi_{mnk} \mathbf{b}_n^T) \right) \\ \Sigma_m^u &= \left[\sum_{n:(m,n) \in \Omega} \sum_{k=1}^K \frac{\tau_{k,1}}{\tau_{k,2}} \phi_{mnk} (\mathbf{b}_n^T \mathbf{b}_n + \Sigma_n^v) + \Lambda \right]^{-1} \end{aligned} \quad (11)$$

在 $\Lambda = \text{diag}(\lambda) - 1$ 。更新为 $\mathbf{B} \cdot$ 相似。我省略由于空间的约束。对于 η , 我们知道

$$\mathbb{E}_q[\lambda_r] = \frac{2b_1 + \mathbf{a}_r^T \mathbf{a}_r + \sum_{m=1}^M (\Sigma_m^u)_{rr} + \mathbf{b}_r^T \mathbf{b}_r + \sum_{n=1}^N (\Sigma_n^v)_{rr}}{2a_1 + M + N} \quad (12)$$

备注：如果我们发现 $E Q [R] \lambda$ 小于一些预定义的阈值，我们会删除相应的尺 R 和 $R 1$ 减量之前。理论基础这背后是，与此列在前一零的意思是，一个非常小的差异表明，本栏目将收缩到零，因此不会有助于解释数据。

通过重复上述步骤的更新，我们放弃了不使用高斯分量和矩阵尺寸而调整自由变分参数的近似原始分布直至收敛。

7. 实验

在本节中，我们会比较建议七个国家的最先进的方法 AMF 1 模型。那里

非贝叶斯方法 (CWM [19]和[5], PCP) 贝叶斯方法 (vblr [1], brpca [6] [28], 多频, B 细胞复制成熟因子[29]和 MOG RPCA (mrpca) [32])。所有的

我们已经进行了实验，的超 AMF 是固定的没有进一步的调整：

$$\bar{a}_0 = \bar{b}_0 = 10^{-4} \bar{a}_1 = \bar{b}_1 = 0.1, \alpha = 1.$$

7. 1. 综合实验

在这一部分中，我们首先遵循 [18]设计三套合成的实验来比较所有性能上述低秩矩阵分解方法。对于这些三组实验中，我们首先将三种不同的类型噪声对地面真实矩阵相应的设置。噪声的细节，如表 1 所示。然后我们滴对每个输入测试这些方法的鲁棒性 20%。对于每个实验我们产生 10 的地面真实低秩矩阵，每一

个都是由 $0 \in R 50 \times 50$ 表示，两个随机生成的低秩矩阵的乘积 $U \in R 50$ 和 $R 50 \times 4 V \in \times 4$ 。我们假设每个元素在 U 和 V 的服从正态分布 $N(0, 1)$ 和地面真实等级是 $R = 4 Y$ 。

	$\mathcal{N}(0, 0.5^2)$	$U[-5, 5]$	$U[-2, 2]$
Gaussian Noise	100%	0	0
Sparse Noise	0	30%	0
Mixture Noise	15%	20%	20%

表 1. 三种类型的噪声。u 表示均匀噪声的范围

回收的低秩矩阵表示

u 和 v 在每一组实验中，我们首先比较这八方法在不同初始等级。考虑到有些方法是不能自动调谐的秩一般，我们初始化 u 和 v 有真实排名 $R = 4$ ，然后我们将两倍的初始等级地面真实等级， $R = 8$ ，由于地面真理等级 1 可用一般在实际应用中的先验未知。对于在 AMF 和高斯分量数 k —主成分分析，我们初步设定为 64 的足够大的交流—容纳各种类型的噪声。具有相同的设置，后来我们滴每个输入 20% 个数据项和重复实验。注意，mrpca, PCP, brpca 和 B 细胞复制成熟因子不是设计来处理丢失的数据，所以我们代替—PCA 与以前版本的 MOG [18]这是人能够处理缺失数据。我们只比较 AMF, MOG, 水煤浆，在这一部分 vblr 和多频。对于性能比较，我们使用的相对误差 Frobenius 范数相对于地面的真理，定义为 $\frac{\|Y_0 - \hat{U}\hat{V}^T\|_F}{\|Y_0\|_F}$ 。每个噪声背景，我们跑 10 次不同的输入 Y_0 和记录他们的表现。平均的相对误差值列于表 2。

从表 2 中，当给出完整的输入，我们可以看到 AMF 优于给所有其他的方法最小相对误差噪声三种不同，即使没有事先的知识是关于组件的数量和正确的秩。即使 20% 的输入项损坏，AMF 显示识别的鲁棒性，相对误差小。作为一个全贝叶斯模型，AMF 达到相当的速度，节约了成本超参数整定和多次试验确定适当数量的集群或矩阵的秩。当高斯噪声类型简单，例如，mrpca, brpca 和 vblr 损坏的输入可实现结果 AMF 的。然而，他们的表现令人瞩目 AMF 在噪声越来越复杂。这个属性 AMF 的高弹性模型未知的复杂噪声并准确地检测等级。

现在我们的重点转移到 AMF 只和经验探讨不同等级 R 及其内在性质初始数量的高斯 K。我们分离的影响这两个参数的变化只是其中一个。我们先把 R = 4 这就是真实值而从 10 到

200 K。然后我们把 K = 64 而改变从 4 到 23。每一组实验进行的基础在相同的地面真实矩阵作为输入。对于每个不同的 R 和 K 的设置，我们用随机 AMF 的五倍初始化。每个均值和标准偏差设置记录在图 3。

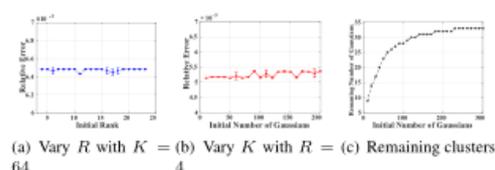


图 3. AMF 本质分析

图 3 (a) 和 3 (b)，AMF 的平均相对误差一几乎看不见最不变的误差棒当一个变化的参数。而在图 3 (c)，高斯的 K 的初始数量是不同的，其它组件的数量总是停留在 33。这两个实验的结果都表明，AMF 相当稳定不管如何初始化和参数集。

Noise	R	AMF	MRPCA/MoG	CWM	PCP	VBLR	BRPCA	BRMF	PRMF
Gaussian	4	0.0451 (0.33)	0.0451 (0.26)	0.0576(0.11)	0.0751(0.57)	0.0758(0.08)	0.0451 (1.54)	0.0545(1.25)	0.0563(0.86)
	8	0.0500 (0.37)	0.0519(1.56)	0.0777(0.15)	0.0914(1.82)	0.0875(0.08)	0.0514(1.90)	0.0747(1.24)	0.0888(0.93)
	4(w/m)	0.0468 (0.45)	0.0517(5.55)	0.0572(0.27)	-	0.0468 (0.15)	-	-	0.0604(0.48)
	8(w/m)	0.0485 (0.49)	0.0847(7.08)	0.0738(0.71)	-	0.0485 (0.16)	-	-	0.0853(0.51)
Sparse	4	6.9E-7 (1.31)	1.9E-6 (0.48)	0.0920(0.23)	0.1342(0.52)	0.0645(0.07)	0.0354(1.31)	0.0320(1.16)	0.0519(0.77)
	8	7.9E-7 (1.51)	3.1E-6(0.61)	0.1880(0.64)	0.1471(1.48)	0.0433(0.07)	0.0797(1.80)	0.1856(1.24)	0.2359(0.89)
	4(w/m)	1.9E-5 (0.87)	0.0513(1.24)	0.0024(0.28)	-	0.3016(0.13)	-	-	0.0273(0.31)
	8(w/m)	2.2E-5 (0.89)	0.2821(4.11)	0.1722(0.87)	-	0.3085(0.17)	-	-	0.2219(0.45)
Mixture	4	0.0050 (1.17)	0.0052(1.13)	0.0754(0.29)	0.1318(0.76)	0.2837(0.11)	0.0096(1.92)	0.0276(1.28)	0.1025(0.23)
	8	0.0062 (1.27)	0.0150(1.60)	0.2152(0.74)	0.1653(1.54)	0.2848(0.14)	0.0442(2.56)	0.1994(1.36)	0.2468(0.24)
	4(w/m)	0.0272 (1.21)	0.0730(1.61)	0.1245(0.63)	-	0.3246(0.18)	-	-	0.1636(0.26)
	8(w/m)	0.0311 (2.11)	0.4427(3.41)	0.3265(0.38)	-	0.3053(0.17)	-	-	0.3413(0.25)

表 2. 平均相对误差八方法三种不同初始等级噪声下。括号中的数字是对应的时间记录 (二) 这些方法。(W / M) 表明丢失的输入情况。最好的结果是以粗体显示。

7.2. 文字去除

接下来我们跟随 [29] 进行文字去除模拟实验。的任务是要删除一些文字嵌入具有一定的图案作为背景图像。这个的干净的图像大小设置为

256 × 256 等级 10 对应的数据矩阵。

图 4 显示输入图像 (图 4 (a)) 而形成的 (背景) 干净的图像 (图 4 (b)) 和 (前景) 异常掩码 (图 4 (c))

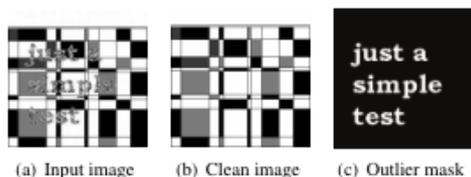


图 4. 文本去除模拟图像

作为真正的排名在现实往往是事先未知一世界的的数据，我们设置的初始最大等级两次被所有算法的真面目。我们更换通过 mbrmf 自 B 细胞复制成熟因子，如 [29] 表明，后者具有更好的处理连续值。图 5 示范文本恢复和重建的背景图像使用不同的算法。

通过视觉检查恢复面具在图 5 中，我们需要注意的是面具和 AMF mbrmf 似乎最尖锐的。经 mrpca 口罩，多频和 PCP 接近。一是可识别的 brpca 但水煤浆和 vblr 效果很模糊。关于对低秩矩阵重构，AMF 给干净的恢复的背景，其次是 mbrmf，PCP 和 mrpca。多频和 brpca 留下太多虽然他们在异常检测异常值做一个公平

的工作。以某种程度上都在执行任务 vblr CWM 失败在恢复的背景略好。

表 3 显示了定量结果。对异常值的检测，虽然 mbrmf 产量最高的 AUC 值，AMF 走得很近。关于背景恢复，AMF 明显比其他低收益的相对误差方法，显示出定量的结果是一致的上述的定性结果。

	AMF	MRPCA	CWM	PCP	VBLR	BRPCA	MBRMF	PRMF
AUC	0.991	0.954	0.867	0.976	0.899	0.921	0.993	0.960
RE	0.068	0.101	0.185	0.103	0.163	0.205	0.102	0.144

表 3. 不同方法的比较。我们使用面积下曲线 (AUC) 测量异常掩码检测和相对对背景恢复的 Frobenius 范数误差。最好的结果以粗体显示。

7.3. 面对阴影去除

在本节中，我们将面临一个现实中的应用研究拍摄的图像在不同的照明。这样的脸阴影去除任务通常是作为一个重要的预处理步骤由许多人脸识别系统。来源的噪音，如低光照、阴影和镜面反射，这项工作带来了巨大的挑战。

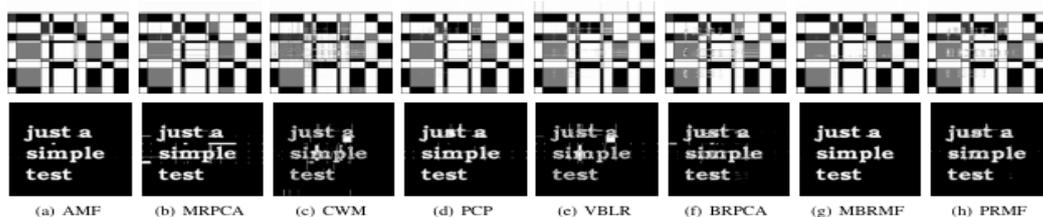


图 5. 通过不同的算法恢复的背景和前景的面具



图 6. 面对阴影的去除效果。每一个从左向右的九组显示原来的面孔和那些回

收的 AMF, mrpca, CWM, PCP, vblr, brpca, mbrmf, 和多频



图 7. 面对阴影去除损坏的输入结果。每个从左到右的六组显示原始的脸
那些由 AMF, MOG, CWM 恢复, vblr 和多频

我们使用扩展 Yale B 两个人形象数据库 [10] 为例。对于每个人来说, 我们使用的所有数据库中的 64 幅图像。自对准人脸图像的同一个人的不同光照下的谎言很接近一个四维线性子空间, 等级设置为 8 所有与 PCP 是异常的方法允许选择等级自动。首先我们使用原来的脸作为输入, 比较八种方法。然后每一个面, 我们随机滴半个像素使用这些损坏的图像作为输入。因为只有 AMF, mrpca, CWN, vblr 和多频都声称能够处理丢失的数据, 我们给出了这五种算法的结果。图 6 7 显示了一些结果。通过观察第一排的图 6, 我们可以看到, 所有的算法在消除显示在正常情况下性能良好一在对夫人的脸上的阴影, 消除眼从人的闪烁。然而, 在极端情况下, 当一个大的脸部的阴影部分, AMF 优于其他通过恢复面部的同时尽可能的方法保留原始面貌。虽然 mrpca 和 brpca 有时可以获得比较的结果, 他们的影响常常是由文物特别是影响男人的形象。大多数的结果 mbrmf 吵无法保存等功能的男人的胡

子。同时在图 7 中, 显示了令人印象深刻的鲁棒性但 AMF 半像素的条目被损坏。相比于其他四种方法, AMF 能够以更少的噪声恢复原来的脸。这些结果再次表明必要的噪声源和分解在这种情况下, AMF 具有巨大的应用潜力。

8. 结论和未来的工作

我们提出了一种新的非参数贝叶斯矩阵分解方法。它解决了两个关键模型选择问题, 通过将 ARD 之前和 DP 之前在系数矩阵和噪声模型, 分别。基于 DP 掰棒子表示, 我们设计了一种高效的变分推理算法。从实验结果中, 我们发现该方法具有较高的潜力, 处理各种具有自动模型选择的应用范围。把这项工作进一步, 在 [34, 29], 我们也要事实上, 在许多经常发生连续的离群挖掘计算机视觉的应用。这意味着一些聚类结构矩阵中的噪声模式。表示此属性在模型明确可能导致进一步的性能的改善。我们将探讨本研究在我们今后的工作方向。

9. 确认

这项研究已经通过研究部分支持
持格兰特 FSGRF14EG36.

References

- [1] S. Babacan, M. Luessi, R. Molina, and A. Katsaggelos. Sparse Bayesian methods for low-rank matrix estimation. *IEEE Transactions on Signal Processing*, 60(8):3964 - 3977, 2012.
- [2] M. Blei and I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian analysis*, 1(1):121 - 143, 2006.
- [3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends R? in Machine Learning*, 3(1):1 - 122, 2011.
- [4] A. M. Buchanan and A. W. Fitzgibbon. Damped Newton algorithms for matrix factorization with missing data. In *CVPR*, pages 316 - 322, 2005.
- [5] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the Association for Computing Machinery*, 58(3), 2011.
- [6] L. Carin, X. Ding, and L. He. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 20(12):3419 - 3430, 2011.
- [7] F. De La Torre and J. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117 - 142, 2003.
- [8] A. Eriksson and A. Van Den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l_1 norm. In *CVPR*, pages 771 - 778, 2010.
- [9] T. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209 - 230, 1973.
- [10] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643 - 660, 2001.
- [11] J. He, L. Balzano, and A. Szlam. Incremental gradient on the

- Grassmannian for online foreground and background separation in subsampled video. In CVPR, pages 1568 - 1575, 2012.
- [12] H. Ishwaran and L. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 2001.
- [13] Q. Ke and T. Kanade. Robust l_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In CVPR, pages 739 - 746, 2005.
- [14] N. D. Lawrence and R. Urtasun. Non-linear matrix factorization with Gaussian processes. In ICML, pages 601 - 608, 2009.
- [15] D. J. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415 - 447, 1992.
- [16] D. J. MacKay. Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3):469 - 505, 1995.
- [17] V. Maz'ya and G. Schmidt. On approximate approximations using Gaussian kernels. *IMA Journal of Numerical Analysis*, 16(1):13 - 29, 1996.
- [18] D. Meng and F. De la Torre. Robust matrix factorization with unknown noise. In ICCV, 2013.
- [19] D. Meng, Z. Xu, L. Zhang, and J. Zhao. A cyclic weighted median method for l_1 low-rank matrix factorization with missing entries. In AAAI, 2013.
- [20] R. M. Neal. Bayesian learning for neural networks. PhD thesis, University of Toronto, 1995.
- [21] J. Pitman. Combinatorial stochastic processes. Technical report, Springer, 2002.
- [22] C. Rasmussen. The infinite Gaussian mixture model. In NIPS, volume 12, pages 554 - 560, 1999.
- [23] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In ICML,

- pages 880 – 887, 2008.
- [24] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. *NIPS*, 20:1257 – 1264, 2008.
- [25] J. Sethuraman. A constructive definition of Dirichlet priors. Technical report, DTIC Document, 1991.
- [26] Y. Shen, Z. Wen, and Y. Zhang. Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optimization Methods and Software*, pages 1 – 25, 2012.
- [27] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137 – 154, 1992.
- [28] N. Wang, T. Yao, J. Wang, and D.-Y. Yeung. A probabilistic approach to robust matrix factorization. In *ECCV*, pages 126 – 139, 2012.
- [29] N. Wang and D.-Y. Yeung. Bayesian robust matrix factorization for image and video processing. In *ICCV*, 2013.
- [30] S. Wang, D. Liu, and Z. Zhang. Nonconvex relaxation approaches to robust matrix recovery. In *IJCAI*, pages 1764 – 1770, 2013.
- [31] J. Xu, V. Ithapu, L. Mukherjee, J. Rehg, and V. Singh. GOSUS: Grassmannian online subspace updates with structured-sparsity. In *ICCV*, 2013.
- [32] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. 2014.
- [33] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust l_1 norm. In *CVPR*, pages 771 – 778, 2012.
- [34] X. Zhou, C. Yang, and W. Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):597 – 610, 2013.
- [35] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma. Face recognition with contiguous occlusion using Markov random fields. In *ICCV*, pages

1050 - 1057, 2009.

