

具有聚类结构相似性的非参数贝叶斯字典学习算法

董道广* 芮国胜 田文彪 张洋 刘歌

(海军航空大学信号与信息处理山东省重点实验室 烟台 264001)

摘要: 利用图像结构信息是字典学习的难点, 针对传统非参数贝叶斯算法对图像结构信息利用不充分, 以及算法运行效率低下的问题, 该文提出一种结构相似性聚类beta过程因子分析(SSC-BPFA)字典学习算法。该算法通过Markov随机场和分层Dirichlet过程实现对图像局部结构相似性和全局聚类差异性的兼顾, 利用变分贝叶斯推断完成对概率模型的高效学习, 在确保算法收敛性的同时具有聚类的自适应性。实验表明, 相比目前非参数贝叶斯字典学习方面的主流算法, 该文算法在图像去噪和插值修复应用中具有更高的表示精度、结构相似性测度和运行效率。

关键词: 变分贝叶斯; Markov随机场; 字典学习; 去噪; 修复

中图分类号: TN911.73; TP301.6

文献标识码: A

文章编号: 1009-5896(2020)11-2765-08

DOI: [10.11999/JEIT190496](https://doi.org/10.11999/JEIT190496)

A Nonparametric Bayesian Dictionary Learning Algorithm with Clustering Structure Similarity

DONG Daoguang RUI Guosheng TIAN Wenbiao ZHANG Yang LIU Ge

(Signal and Information Processing Key Laboratory in Shandong,
Navy Aviation University, Yantai 264001, China)

Abstract: Making use of image structure information is a difficult problem in dictionary learning, the traditional nonparametric Bayesian algorithms lack the ability to make full use of image structure information, and faces problem of inefficiency. To this end, a dictionary learning algorithm called Structure Similarity Clustering-Beta Process Factor Analysis (SSC-BPFA) is proposed in this paper, which completes efficient learning of the probabilistic model via variational Bayesian inference and ensures the convergence and self-adaptability of the algorithm. Image denoising and inpainting experiments show that this algorithm has significant advantages in representation accuracy, structure similarity index and running efficiency compared with the existing nonparametric Bayesian dictionary learning algorithms.

Key words: Variational Bayesian; Markov Random Field (MRF); Dictionary Learning (DL); Denoising; Inpainting

1 引言

近年来, 非参数贝叶斯(Non-Parametric Bayesian, NPB)学习^[1-4]逐步从数据分类^[5,6]迁移到字典学习^[7-12](Dictionary Learning, DL)的应用中, 相比K-SVD^[1]等传统的综合DL算法, 它凭借能够自适应地推断未知噪声水平和字典规模的优势^[9]而受到关注。早前, Zhou等人^[9]利用Knowles等人^[13]提出的Beta过程因子分析(Beta Process Factor

Analysis, BPFA)模型对DL问题进行NPB概率建模, 提出了NPB-DL领域最为主流且经典的BPFA算法, 并在分类^[14,15]、压缩^[2]、去噪^[9,10]、修复^[11]和超分辨^[16]等领域取得成功应用。尽管如此, 现有的NPB-DL算法普遍将训练集内的数据样本建模为彼此独立的存在, 样本间的相关性所蕴含的数据结构信息则遭到忽视, 这制约了NPB-DL所得字典对数据拟合能力的进一步提升。

凭借出色的聚类/分类能力, NPB天然地适合对数据中存在的结构相似性进行挖掘, Zhou等人^[9]后续提出的DP-BPFA算法和PSBP-BPFA算法是这方面工作的典型代表。DP-BPFA算法通过Dirichlet过程进行数据的空域聚类, 从而实现了对数据结构的非局部相似性的挖掘, 但忽视了数据结构的局部

收稿日期: 2019-07-03; 改回日期: 2020-02-28; 网络出版: 2020-09-01

*通信作者: 董道广 sikongyu@yeah.net

基金项目: 国家自然科学基金(41606117, 41476089, 61671016)

Foundation Items: The National Natural Science Foundation of China (41606117, 41476089, 61671016)

相似性,无法保证其聚类结果的空间平滑性。PSBP-BPFA算法将数据的空间位置信息引入高斯核函数,进而构建多项Probit回归模型,并将其以累积概率密度函数的形式作用于聚类标签先验概率的折棍构造(stick-breaking construction),从而将结构的局部相似性约束引入数据的空域聚类中,一定程度上兼顾了聚类的空间平滑性。该算法的概率模型不具备共轭性和解析闭式形式,所以无法使用变分贝叶斯(Variational Bayesian, VB)方法进行高效的确定性推断,而传统的采样策略对回归系数进行推断学习则要付出极大的计算开支,此外该算法缺乏对核宽度的选择机制,如果为核宽度额外增加一层概率模型,则会继续加重算法的计算负担。笔者前期曾提出一种基于结构相似性(Structure Similarity, SSIM)的NPB-DL算法,即SSIM-BPFA算法^[7],在BPFA算法的框架基础上对数据样本进行聚类预处理,但自适应性较差,需要预先人为设定聚类数目。

综上,为了更好地兼顾NPB-DL对非局部和局部相似性的利用,并提高算法的运行效率和自适应性,本文提出一种结构相似性聚类beta过程因子分析(Structure Similarity Clustering-Beta Process Factor Analysis, SSC-BPFA)字典学习算法。首先,在聚类标签的先验概率建模中引入Markov随机场(Markov Random Fields, MRF),从而为NPB-DL施加数据结构的局部相似性约束;其次,为聚类标签的先验概率施加Dirichlet先验分布并利用所构建的MRF对其进行参数化,确保了概率模型的解析性和共轭性;最后,采用VB对模型进行推断学习,自适应地推断出合理的聚类数目,理论上能够收敛到全局最优,并提高了算法的运行效率。图像去噪和修复实验结果表明,相比K-SVD和前述几种NPB-DL算法,本文算法具有更高的信误比(Signal-to-Error Ratio, SER)、结构相似性测度(Structure SIMilarity index, SSIM)和运行效率。

2 基于聚类结构相似性的NPB概率建模

以灰度图像的处理为例,彩色图像的处理方式可以类似地拓展获得。将图像中的全部 $M \times M$ 像素块经过列堆栈处理所得的 N 个 P 维向量构成训练集 $\{\mathbf{x}_i\}_{i=1}^N$ (其中 $P = M \times M$)。将字典内原子 \mathbf{d}_k 的未知数目 K 截断为一个足够大的正整数,通常取 $K = 256$ 或 512 ,字典记作 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in \mathbb{R}^{P \times K}$,则任意数据样本 \mathbf{x}_i 在字典下可表示为式(1),其中 $\mathbf{s}_i \in \mathbb{R}^{K \times 1}$ 是权重向量,0-1二值向量 $\mathbf{z}_i \in \mathbb{R}^{K \times 1}$ 是稀疏模式指示向量, \odot 表示逐元素乘积, $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{P \times 1}$ 表示残差。稀疏模式指示向量 \mathbf{z}_i 服从由原

子使用概率 $\boldsymbol{\pi}_i \in \mathbb{R}^{K \times 1}$ 参数化的Bernoulli分布,如式(2)所示,其中 z_{ik} 和 π_{ik} 分别是 \mathbf{z}_i 和 $\boldsymbol{\pi}_i$ 的第 k 个元素。

$$\mathbf{x}_i = \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i) + \boldsymbol{\varepsilon}_i \quad (1)$$

$$\mathbf{z}_i \stackrel{\text{i.i.d.}}{\sim} \prod_{k=1}^K \text{Bernoulli}(z_{ik} | \pi_{ik}) \quad (2)$$

本节的NPB概率模型对训练集内的全部数据样本进行空域聚类,以实现对非局部相似性结构的利用,将聚类数目 L 的初始值截断为某个较大的正整数 L_{\max} ,相同聚类归属的样本具有相同的原子使用概率。这个过程可以表达为式(3)的Dirichlet过程(DP),其中 G_i 是样本 \mathbf{x}_i 的聚类先验测度, G_0 是基底测度, α_0 是尺度参数,按照Dirichlet过程的离散性质,该式又可进一步具体地表达为式(4)的点测度离散加权形式,其中原子使用概率 π_l^* 对应于某个点测度,是从基底测度 G_0 中采样获得的,混合权重 β_{il} 是尺度参数 α_0 的函数,代表样本 \mathbf{x}_i 属于聚类 l 并且原子使用概率为 $\boldsymbol{\pi}_i = \boldsymbol{\pi}_l^*$ 。采用式(5)的Beta分布作为测度 G_0 的具体形式,其中 $a_0 = b_0 = 1$, π_{ik}^* 是 $\boldsymbol{\pi}_i^*$ 的第 k 个元素, α_0 取为 $1/L$ 。显然,混合权重 $\beta_{i1}, \beta_{i2}, \dots, \beta_{iL}$ 应该满足加和为1,因此本文为 $\beta_{i1}, \beta_{i2}, \dots, \beta_{iL}$ 赋予式(6)所示的Dirichlet先验分布,其中参数 $\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{iL}$ 均为 α_0 的函数。

$$G_i \sim \text{DP}(\alpha_0, G_0) \quad (3)$$

$$G_i = \sum_{l=1}^L \beta_{il} \cdot \delta(\boldsymbol{\pi}_i - \boldsymbol{\pi}_l^*) \quad (4)$$

$$\boldsymbol{\pi}_i^* \stackrel{\text{i.i.d.}}{\sim} \prod_{k=1}^K \text{Beta}(\pi_{ik}^* | a_0/K, b_0(K-1)/K) \quad (5)$$

$$\{\beta_{il}\}_{l=1}^L \sim \text{Dirichlet}(\beta_{i1}, \beta_{i2}, \dots, \beta_{iL} | \lambda_{i1}, \lambda_{i2}, \dots, \lambda_{iL}) \quad (6)$$

在式(3)–式(6)的Dirichlet混合模型基础上,本文为每个 \mathbf{x}_i 赋予一个聚类标签 $t_i \in \{1, 2, \dots, L\}$,并假设 t_i 服从式(7)的多项分布,相应的分布参数 $\beta_{i1}, \beta_{i2}, \dots, \beta_{iL}$ 正是该模型的混合权重。与 \mathbf{x}_i 相应的 $\boldsymbol{\pi}_i$ 则可表达为式(8),从中可见无论训练集内样本数目有多大, $\boldsymbol{\pi}_i$ 的取值数目总是有限的,事实上随着最终聚类数目的自适应推断确定,需要推断学习的 $\boldsymbol{\pi}_i$ 数目将显著小于 L 。

$$t_i \stackrel{\text{i.i.d.}}{\sim} \text{Multi}(t_i = 1, t_i = 2, \dots, t_i = L | \beta_{i1}, \beta_{i2}, \dots, \beta_{iL}) \quad (7)$$

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_{t_i}^*, t_i = l, l \in \{1, 2, \dots, L\} \quad (8)$$

上述基于聚类实现了对数据结构非局部相似性的利用,为了兼顾局部相似性,本文在聚类标签的先验概率分布中建立MRF,以提升聚类的空间平

滑性。具体方法是, 以任意数据样本 \mathbf{x}_i 对应的像素块为中心, 取其 3×3 个邻域像素块对应的样本构成 \mathbf{x}_i 的邻域数据集合 Ω_i , 将该集合内样本的聚类标签进行空域平滑处理获得MRF因子 G_{il} , 如式(9)所示, 随后将混合权重 $\beta_{i1}, \beta_{i2}, \dots, \beta_{iL}$ 的先验参数 $\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{iL}$ 表达为 α_0 与 $G_{i1}, G_{i2}, \dots, G_{iL}$ 的乘积, 如式(10)所示。

$$G_{il} = \frac{1}{|\Omega_i|} \sum_{\mathbf{x}_m \in \Omega_i} I(t_m = l) \quad (9)$$

$$\lambda_{il} = \alpha_0 G_{il} \quad (10)$$

G_{il} 值越大表明聚类 l 在邻域数据系统中越“热门”, 相应的混合权重 β_{il} 就越大, 样本 \mathbf{x}_i 属于该热门聚类的概率就越高, 这使得数据样本以更高的概率与其大多数邻域样本同属一个聚类, 因而体现出了结构的局部相似性, 提高了聚类结果的空间平滑性。与PSBP-BPFA算法相比, 上述概率模型的首要优点是具有共轭性和解析性, 可以通过VB方法进行更为高效的确定性推断; 其次, 通过对邻域聚类标签平滑的方式构建MRF, 使施加在聚类先验概率上的局部相似性约束作用仅由邻域范围内的少量样本产生, 无需像PSBP-BPFA那样对全部样本逐个计算核函数值, 同样有利于节省计算开支。

式(1)中的原子 \mathbf{d}_k 、表示权重 \mathbf{s}_i 及残差 $\boldsymbol{\varepsilon}_i$ 则沿用BPFA框架提供的先验分布, 如式(11)–式(15)所示, 其中 γ_s 和 γ_ε 分别是表示权重和残差的精度参数, c_0, d_0, e_0 和 f_0 是Gamma超参数, 为保证先验的非信息性, 均取 10^{-6} 。

$$\mathbf{d}_k \sim \mathcal{N}(\mathbf{d}_k | \mathbf{0}, P^{-1} \mathbf{I}_P) \quad (11)$$

$$\mathbf{s}_i \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(\mathbf{s}_i | \mathbf{0}, \gamma_s^{-1} \mathbf{I}_K) \quad (12)$$

$$\gamma_s \sim \text{Gamma}(\gamma_s | c_0, d_0) \quad (13)$$

$$\boldsymbol{\varepsilon}_i \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(\boldsymbol{\varepsilon}_i | \mathbf{0}, \gamma_\varepsilon^{-1} \mathbf{I}_P) \quad (14)$$

$$\gamma_\varepsilon \sim \text{Gamma}(\gamma_\varepsilon | e_0, f_0) \quad (15)$$

综上所述, 式(1)–式(15)构成了本文算法的NPB概率模型, 相应的图模型如图1所示。

3 算法总结及分析

在上节搭建的概率模型中, 观测数据是 $\{\mathbf{x}_i\}_{i=1}^N$, 隐变量集是 $\Theta = \{\{\mathbf{d}_k\}_{k=1}^K, \{\mathbf{s}_i, z_{ik}\}_{i=1, k=1}^{N, K}, \{\pi_{ik}^*\}_{i=1, k=1}^{N, K}, \{t_i\}_{i=1}^N, \{\beta_{il}\}_{i=1, l=1}^{N, L}, \gamma_s, \gamma_\varepsilon\}$, 超参数集为 $\Xi = \{a_0, b_0, c_0, d_0, e_0, f_0, \alpha_0\}$ 。本节采用VB方法对隐变量进行推断学习。VB推断的理论基础可以参考文献[18], 本文不再赘述, 只将上节概率模型框架内每个隐变量的变分分布及其VB推断的更新公式的推导结果列于表1, 其中变分参数集为 $\mathbf{H} =$

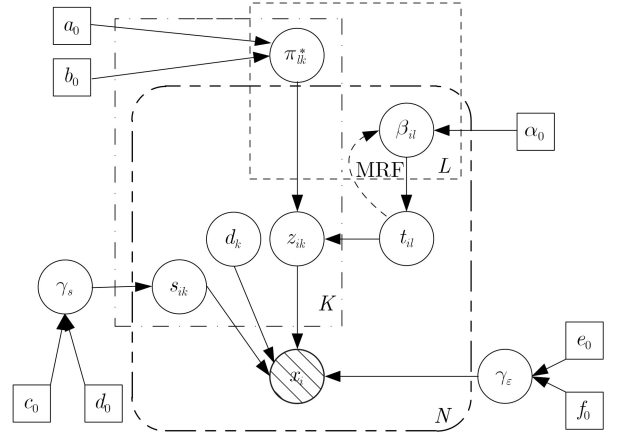


图1 本文算法的概率图模型

$\{\{\tilde{\mathbf{m}}_k, \tilde{\boldsymbol{\Lambda}}_k\}_{k=1}^K, \{\tilde{\boldsymbol{\mu}}_{ik}, \tilde{\sigma}_{ik}^2, \tilde{\rho}_{ik}\}_{i=1, k=1}^{N, K}, \{\tilde{a}_{lk}, \tilde{b}_{lk}\}_{l=1, k=1}^{L, K}, \{\tilde{\xi}_{il}, \tilde{\lambda}_{il}\}_{i=1, l=1}^{N, L}, \tilde{c}_0, \tilde{d}_0, \tilde{e}_0, \tilde{f}_0\}$ 。隐变量 t_i 和 β_{il} 及其变分参数 $\tilde{\xi}_{il}$ 和 $\tilde{\lambda}_{il}$ 的初始化通过对所有样本 $\{\mathbf{x}_i\}_{i=1}^N$ 进行K均值聚类实现, 其余隐变量及其超参数的初始化与BPFA算法相同。本文算法的详细步骤可参考算法1, 这里将该算法称为结构相似性聚类BPFA (Structure Similarity Clustering-BPFA, SSC-BPFA) 算法。

由于该算法的概率模型满足共轭性, 因此模型中每个隐变量恰好具有相同形式的变分分布和真实后验分布, 所以VB推断实际上退化为期望最大化估计, 后者可以保证未知参数的估计收敛到全局最优, 这方面有坚实的理论研究基础作为支撑, 感兴趣的读者可以参考文献[18], 而一旦模型的共轭性被破坏, 那么这种全局最优将不复存在, 这正是本文追求概率建模共轭性的原因所在。

4 数值实验及结果分析

使用House, Barbara和Lena等经典图像作为实例, 在硬件条件为2.4GHz CPU和8G内存的计算机上对本文算法开展数值实验。实验从字典学习算法的图像去噪和插值修复应用方面展开, 参与比较的算法包括BPFA, DP-BPFA, PSBP-BPFA, SSIM-BPFA以及K-SVD算法。采用信误比(Signal-to-Error Ratio, SER)和结构相似性测度(Structural SIMilarity index, SSIM)这两种指标来衡量字典的应用性能, 同时将平均运行时间用于衡量算法的运行效率。

在进行对比实验前, 先给出本文算法分别对Peppers, Barbara, Lena, House, Cameraman和Boat 6幅图像进行字典学习过程中获得的聚类效果, 如图2所示。该聚类即是基于原子使用概率的差别而实现的, 并将同一聚类的图像块取为相同的像素值获得, 其中 L 是本文算法推断获得的聚类数

目, 尽管图2的聚类效果只是本文算法的副产品, 但与SSIM-BPFA算法那种人为预设聚类数目并执行完K-均值聚类预处理后才开始进行字典学习的方式相比, 本文算法在挖掘图像结构相似性信息方面的自适应性确实获得有效提高。

图3是不同算法的部分图像去噪实验结果, 其中Boat和Cameraman分别被加入5 dB信噪比的高斯噪声, 尽管去噪后的视觉效果差异并不太显著, 但从图像稀疏表示的SER值来看, 本文算法字典的图像稀疏表示精度高于其他算法。为了验证不同噪

表1 模型中隐变量及其变分参数的VB推断更新公式

隐变量	变分分布	变分参数的VB推断更新公式	隐变量更新公式
\mathbf{d}_k	$q(\mathbf{d}_k) \propto \text{Normal}(\mathbf{d}_k \tilde{\mathbf{m}}_k, \tilde{\Lambda}_k)$	$\tilde{\Lambda}_k = P \mathbf{I}_P + \sum_{i=1}^N \frac{\tilde{e}_0}{\tilde{f}_0} (\tilde{\mu}_{ik}^2 + \tilde{\sigma}_{ik}^2) \tilde{\rho}_{ik} \mathbf{I}_P,$ $\tilde{\mathbf{m}}_k = \tilde{\Lambda}_k^{-1} \cdot \left(\sum_{i=1}^N \frac{\tilde{e}_0}{\tilde{f}_0} \tilde{\mu}_{ik} \tilde{\rho}_{ik} \mathbf{x}_i^{-k} \right)$	$\mathbf{d}_k = \tilde{\mathbf{m}}_k$
s_{ik}	$q(s_{ik}) \propto \text{Normal}(s_{ik} \tilde{\mu}_{ik}, \tilde{\sigma}_{ik}^2)$	$\tilde{\sigma}_{ik}^2 = \left\{ \frac{\tilde{e}_0}{\tilde{f}_0} \tilde{\rho}_{ik} \left[\tilde{\mathbf{m}}_k^T \tilde{\mathbf{m}}_k + \text{tr}(\tilde{\Lambda}_k^{-1}) \right] + \frac{\tilde{c}_0}{\tilde{f}_0} \right\}^{-1},$ $\tilde{\mu}_{ik} = \tilde{\sigma}_{ik}^2 \cdot \left(\frac{\tilde{e}_0}{\tilde{f}_0} \tilde{\rho}_{ik} \tilde{\mathbf{m}}_k^T \mathbf{x}_i^{-k} \right)$	$s_{ik} = \tilde{\mu}_{ik}$
		$\tilde{\rho}_{ik} = \frac{\rho_{ik,1}}{\rho_{ik,1} + \rho_{ik,0}},$	
z_{ik}	$q(z_{ik}) \propto \text{Bernoulli}(z_{ik} \tilde{\rho}_{ik})$	$\rho_{ik,0} = \exp \left\{ \sum_{l=1}^L \tilde{\xi}_{il} \cdot [\psi(\tilde{b}_{lk}) - \psi(\tilde{a}_{lk} + \tilde{b}_{lk})] \right\}$ $\rho_{ik,1} = \exp \left\{ -\frac{1}{2} \frac{\tilde{e}_0}{\tilde{f}_0} (\tilde{\mu}_{ik}^2 + \tilde{\sigma}_{ik}^2) (\tilde{\mathbf{m}}_k^T \tilde{\mathbf{m}}_k + \text{tr}(\tilde{\Lambda}_k^{-1})) \right.$ $\left. + \frac{\tilde{e}_0}{\tilde{f}_0} \tilde{\mu}_{ik} \tilde{\mathbf{m}}_k^T \mathbf{x}_i^{-k} + \sum_{l=1}^L \tilde{\xi}_{il} \cdot [\psi(\tilde{a}_{lk}) - \psi(\tilde{a}_{lk} + \tilde{b}_{lk})] \right\}$	$z_{ik} = \tilde{\rho}_{ik}$
π_{lk}^*	$q(\pi_{lk}^*) \propto \text{Beta}(\pi_{lk}^* \tilde{a}_{lk}, \tilde{b}_{lk})$	$\tilde{a}_{lk} = a_0/K + \sum_{i=1}^N \tilde{\rho}_{ik} \tilde{\xi}_{il}, \tilde{b}_{lk} = b_0(K-1)/K + \sum_{i=1}^N (1 - \tilde{\rho}_{ik}) \tilde{\xi}_{il}$ $\tilde{\xi}_{il} = \xi_{il} / \sum_{l'=1}^L \xi_{il'}$	$\pi_{lk}^* = \frac{\tilde{a}_{lk}}{\tilde{a}_{lk} + \tilde{b}_{lk}}$
t_i	$q(t_i) \propto \text{Multi}(\tilde{\xi}_{i1}, \tilde{\xi}_{i2}, \dots, \tilde{\xi}_{iL})$	$\xi_{il} = \exp \left\{ \sum_{k=1}^K \tilde{\rho}_{ik} [\psi(\tilde{a}_{lk}) - \psi(\tilde{a}_{lk} + \tilde{b}_{lk})] \right.$ $\left. + \sum_{k=1}^K (1 - \tilde{\rho}_{ik}) [\psi(\tilde{b}_{lk}) - \psi(\tilde{a}_{lk} + \tilde{b}_{lk})] \right.$ $\left. + \psi(\tilde{\xi}_{il}) - \psi \left(\sum_{l'=1}^L \tilde{\xi}_{il'} \right) \right\}$	$t_i = \arg \max_l \left\{ \tilde{\xi}_{il}, \forall l \in [1, L] \right\}$
β_{il}	$q(\{\beta_{il}\}_{i=1}^L) \propto \text{Diri}(\tilde{\lambda}_{i1}, \tilde{\lambda}_{i2}, \dots, \tilde{\lambda}_{iL})$	$\tilde{\lambda}_{il} = \tilde{\xi}_{il} + \alpha_0 G_{il}$	$\beta_{il} = \tilde{\lambda}_{il} / \sum_{l'=1}^L \tilde{\lambda}_{il'}$
γ_s	$q(\gamma_s) \propto \text{Gamma}(\gamma_s \tilde{c}_0, \tilde{d}_0)$	$\tilde{c}_0 = c_0 + \frac{1}{2} NK, \tilde{d}_0 = d_0 + \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K (\tilde{\mu}_{ik}^2 + \tilde{\sigma}_{ik}^2)$	$\gamma_s = \tilde{c}_0 / \tilde{d}_0$
γ_ε	$q(\gamma_\varepsilon) \propto \text{Gamma}(\gamma_\varepsilon \tilde{e}_0, \tilde{f}_0)$	$\tilde{e}_0 = e_0 + \frac{1}{2} NP, \tilde{f}_0 = f_0 + \frac{1}{2} \sum_{i=1}^N \left\ \mathbf{x}_i - \hat{\mathbf{D}}(\hat{\mathbf{s}}_i \odot \hat{\mathbf{z}}_i) \right\ ^2$	$\gamma_\varepsilon = \tilde{e}_0 / \tilde{f}_0$

算法1 SSC-BPFA算法

输入: 训练数据样本集 $\{\mathbf{x}_i\}_{i=1}^N$

输出: 字典 \mathbf{D} 、权重 $\{\mathbf{s}_i\}_{i=1}^N$ 、稀疏模式指示向量 $\{\mathbf{z}_i\}_{i=1}^N$ 及聚类标签 $\{t_i\}_{i=1}^N$

步骤1 设置收敛阈值 τ 和迭代次数上限 itr_{\max} , 初始化原子数目 K 与聚类数目 L , 通过K-均值算法对数据样本进行初始聚类;

步骤2 按照式(1)–式(15)完成NPB-DL的概率建模;

步骤3 初始化隐变量集 Θ 和变分参数集 \mathbf{H} , 计算 $\hat{\mathbf{X}} = \mathbf{D}(\mathbf{S} \odot \mathbf{Z})$, 令迭代索引 $\text{itr}=1$;

步骤4 按照表1的VB推断公式对 Θ 和 \mathbf{H} 进行更新, 计算 $\hat{\mathbf{X}}^{\text{new}} = \mathbf{D}(\mathbf{S} \odot \mathbf{Z})$;

步骤5 令 itr 值增加1, 删除 \mathbf{D} 中未使用的原子并更新 K 的值, 计算 $r = \left\| \hat{\mathbf{X}}^{\text{new}} - \hat{\mathbf{X}} \right\|_{\text{F}}^2 / \left\| \hat{\mathbf{X}} \right\|_{\text{F}}^2$, 若 $r < \tau$ 或 $\text{itr} \geq \text{itr}_{\max}$, 删除所含样本占全部样本数量比例低于 10^{-3} 的聚类, 将被删聚类内的样本分配到剩余聚类中 $\tilde{\xi}_{il}$ 最大的那个聚类中, 进入步骤6, 否则跳回步骤4;

步骤6 固定 t_i 和 β_{il} 及其变分参数的估计结果, 将 $\mathbf{d}_k, s_{ik}, z_{ik}, \pi_{lk}^*, \gamma_s$ 和 γ_ε 这6个隐变量及其对应的变分参数继续进行迭代优化更新, 直至重新达到收敛。

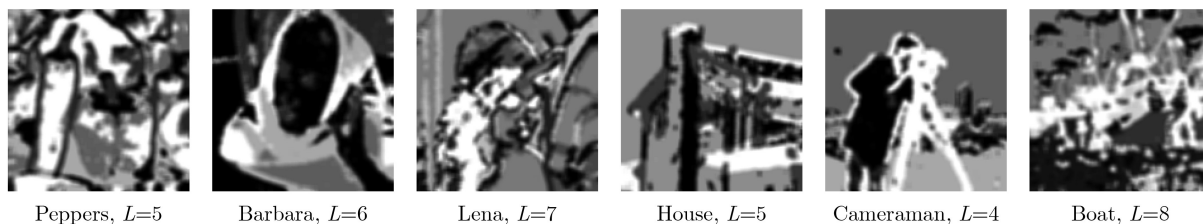


图2 本文算法字典学习过程中获得的聚类效果

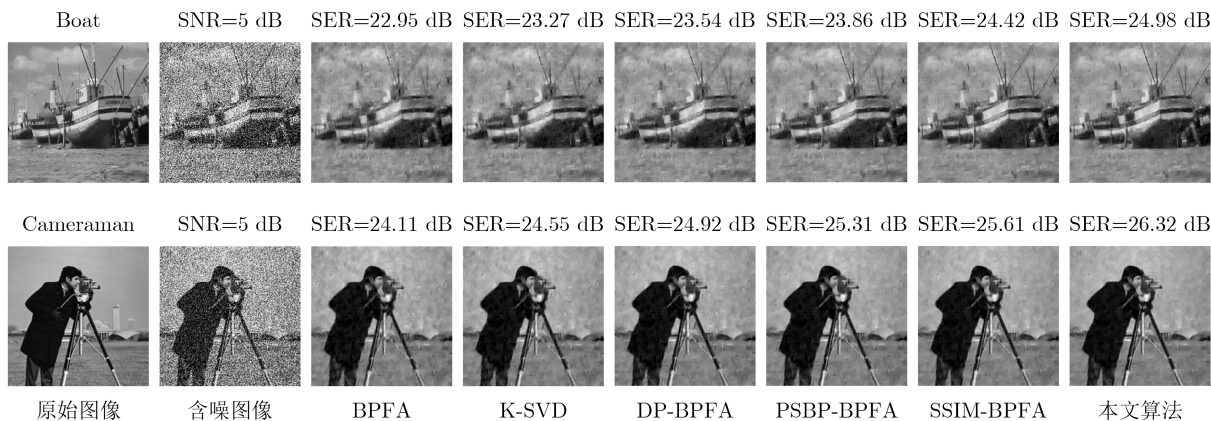


图3 不同算法的图像去噪效果

声强度下的算法去噪性能，图4给出了信噪比从5 dB到30 dB条件下不同算法对Peppers, Barbara, Lena, House, Cameraman和Boat进行去噪所得的平均SER结果，该结果是50次独立实验后计算平均值获得的。

从图4可见，6种算法的总体去噪SER从低到高依次是BPFA算法、K-SVD算法、DP-BPFA算法、PSBP-BPFA算法、SSIM-BPFA算法和本文算法，其中K-SVD与BPFA差距很小，它们都没有对数据的结构加以利用；DP-BPFA和PSBP-BPFA的差距也比较小，它们利用了数据中的结构信息，但如引言所述，利用得不够充分；SSIM-BPFA算法在数据聚类预处理的基础上引入了稀疏表示的块结构特性，其对数据的拟合精度较前4种算法都有显著提高，本文算法相对SSIM-BPFA算法的优势在于自适应性更强，聚类是在字典学习的过程中自动推断进行的，聚类的效果与字典学习更为匹配，因此体现出了去噪方面的精度优势。由于去噪应用中信噪比对SSIM值和算法运行时间的影响不显著，特别是几种算法的SSIM值均接近1，因此这两方面的实验结果将在图像修复实验中体现，以观察数据缺失率对它们的影响。

图5所示是80%的随机像素缺失条件下部分图像的插值修复实验结果，其中第1列是像素缺失后的受损图像，剩余的各列从左向右依次是BPFA, K-SVD, DP-BPFA, PSBP-BPFA, SSIM-BPFA和

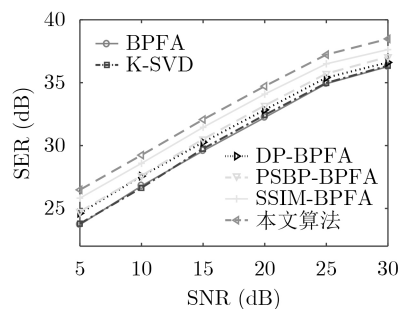


图4 不同信噪比条件下的去噪信噪比

本文算法的修复结果，总体上本文算法的SER值高于其他算法，且视觉效果相对而言也是最好的。图6和图7分别给出了图像数据缺失率从10%到90%的条件下，几种算法对6幅测试图像进行修复的平均SER和SSIM结果，均为50次独立实验后计算平均值所得。

在图6中，随着数据缺失率的增加，所有算法的SER都发生下降，K-SVD算法在低缺失率条件下表现出了相对BPFA和DP-BPFA算法的优势，而在高缺失率条件下K-SVD算法的SER是所有算法中最低的，这说明NPB-DL算法比K-SVD算法的鲁棒性强，几种NPB-DL算法按照SER从低到高排序依次是BPFA, DP-BPFA, PSBP-BPFA, SSIM-BPFA和本文算法，其中本文算法凭借更好的自适应性在图像插值修复应用中体现出相对SSIM-BPFA算法的精度优势。

在图7中，所有算法的SSIM值均随缺失率的增



图5 不同算法的图像修复效果

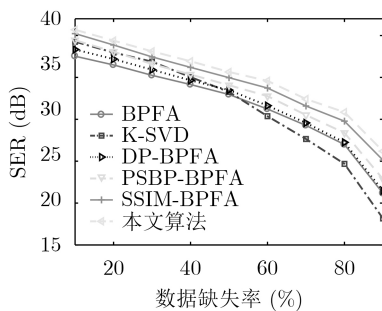


图6 不同缺失率条件下的平均SER结果

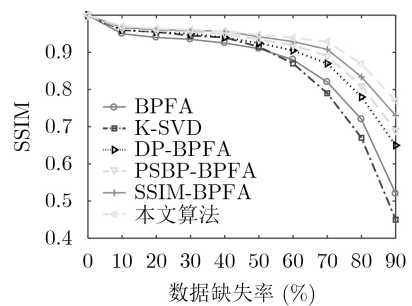


图7 不同缺失率条件下的平均SSIM结果

加而减小，在低缺失率条件下，几种算法的SSIM差距很小，随着缺失率的增加，不同算法SSIM值减小的速度差距显著地拉开，其中按照SSIM从低到高依次是K-SVD, BPFA, DP-BPFA, PSBP-BPFA, SSIM-BPFA和本文算法，这表明本文算法具有对数据结构更好的拟合能力，因而经它修复后图像的SSIM指标也更高。

图8所示是不同数据缺失率条件下几种算法运行所付出的平均时间代价，从中可见随着缺失率的增加，几种算法的平均时间代价都发生减小，PSBP-BPFA算法因其复杂的模型推断机制付出的平均时间代价最高，本文算法由于采用了VB推断机制，避免了计算开支较大的采样操作，因而体现出明显的运行效率优势。

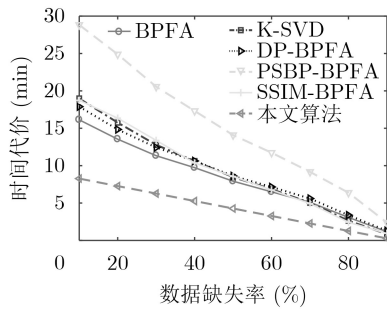


图 8 不同缺失率条件下的平均时间代价

5 结论

本文提出了一种基于聚类结构相似性的SSC-BPFA算法, 该算法将Markov随机场与Dirichlet过程聚类相结合, 从而在利用数据结构非局部相似性的同时兼顾了对结构局部相似性的挖掘, 使聚类的结果在图像空域上更加平滑。同时, 在算法的非参数贝叶斯概率建模过程中确保了模型的共轭性, 从而能够使用变分贝叶斯推断进行模型参数的高效学习, 相比传统NPB-DL算法普遍采用的采样推断方式, 获得了运行效率的显著提升。与NPB-DL领域现有的主流算法以及K-SVD算法相比, 本文算法在图像去噪和插值修复应用中具有更高的信误比、结构相似性测度和更少的运行时间代价。下一步将研究NPB-DL算法对椒盐、泊松等非高斯噪声的去噪性能, 并针对像素非随机缺失情形的图像修复问题进行研究。

参考文献

- [1] XUAN Junyu, LU Jie, ZHANG Guangquan, *et al.* Doubly nonparametric sparse nonnegative matrix factorization based on dependent Indian buffet processes[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(5): 1835–1849. doi: [10.1109/TNNLS.2017.2676817](https://doi.org/10.1109/TNNLS.2017.2676817).
- [2] LI Shaoyang, TAO Xiaoming, and LU Jianhua. Variational Bayesian inference for nonparametric signal compressive sensing on structured manifolds[C]. 2017 IEEE International Conference on Communications, Paris, France, 2017. doi: [10.1109/ICC.2017.7996389](https://doi.org/10.1109/ICC.2017.7996389).
- [3] DANG H P and CHAINAIS P. Indian buffet process dictionary learning: Algorithms and applications to image processing[J]. *International Journal of Approximate Reasoning*, 2017, 83: 1–20. doi: [10.1016/j.ijar.2016.12.010](https://doi.org/10.1016/j.ijar.2016.12.010).
- [4] SERRA J G, TESTA M, MOLINA R, *et al.* Bayesian K-SVD using fast variational inference[J]. *IEEE Transactions on Image Processing*, 2017, 26(7): 3344–3359. doi: [10.1109/TIP.2017.2681436](https://doi.org/10.1109/TIP.2017.2681436).
- [5] NGUYEN V, PHUNG D, BUI H, *et al.* Discriminative Bayesian nonparametric clustering[C]. The 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017. doi: [10.24963/ijcai.2017/355](https://doi.org/10.24963/ijcai.2017/355).
- [6] HUYNH V and PHUNG D. Streaming clustering with Bayesian nonparametric models[J]. *Neurocomputing*, 2017, 258: 52–62. doi: [10.1016/j.neucom.2017.02.078](https://doi.org/10.1016/j.neucom.2017.02.078).
- [7] AHARON M, ELAD M, and BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322. doi: [10.1109/TSP.2006.881199](https://doi.org/10.1109/TSP.2006.881199).
- [8] ELAD M and AHARON M. Image denoising via sparse and redundant representations over learned dictionaries[J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 3736–3745. doi: [10.1109/TIP.2006.881969](https://doi.org/10.1109/TIP.2006.881969).
- [9] ZHOU Mingyuan, CHEN Haojun, PAISLEY J, *et al.* Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images[J]. *IEEE Transactions on Image Processing*, 2012, 21(1): 130–144. doi: [10.1109/TIP.2011.2160072](https://doi.org/10.1109/TIP.2011.2160072).
- [10] ZHOU Mingyuan, PAISLEY J, and CARIN L. Nonparametric learning of dictionaries for sparse representation of sensor signals[C]. The 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Aruba, Netherlands, 2009. doi: [10.1109/CAMSAP.2009.5413290](https://doi.org/10.1109/CAMSAP.2009.5413290).
- [11] PAISLEY J, ZHOU Mingyuan, SAPIRO G, *et al.* Nonparametric image interpolation and dictionary learning using spatially-dependent Dirichlet and beta process priors[C]. 2010 IEEE International Conference on Image Processing, Hong Kong, China, 2010. doi: [10.1109/ICIP.2010.5653350](https://doi.org/10.1109/ICIP.2010.5653350).
- [12] JU Fujiao, SUN Yanfeng, GAO Junbin, *et al.* Nonparametric tensor dictionary learning with beta process priors[J]. *Neurocomputing*, 2016, 218: 120–130. doi: [10.1016/j.neucom.2016.08.064](https://doi.org/10.1016/j.neucom.2016.08.064).
- [13] KNOWLES D and GHARAMANI Z. Infinite sparse factor analysis and infinite independent components analysis[C]. The 7th International Conference on Independent Component Analysis and Signal Separation, London, UK, 2007: 381–388. doi: [10.1007/978-3-540-74494-8_48](https://doi.org/10.1007/978-3-540-74494-8_48).
- [14] ZHANG Linlin, GUINDANI M, VERSACE F, *et al.* A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses[J]. *NeuroImage*, 2014, 95: 162–175. doi: [10.1016/j.neuroimage.2014.03.024](https://doi.org/10.1016/j.neuroimage.2014.03.024).
- [15] AKHTAR N and MIAN A. Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(9): 4038–4050. doi: [10.1109/TNNLS.2017.2742528](https://doi.org/10.1109/TNNLS.2017.2742528).
- [16] POLATKAN G, ZHOU Mingyuan, CARIN L, *et al.* A

- Bayesian nonparametric approach to image super-resolution[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(2): 346–358. doi: [10.1109/TPAMI.2014.2321404](https://doi.org/10.1109/TPAMI.2014.2321404).
- [17] 董道广, 芮国胜, 田文飏, 等. 基于结构相似性的非参数贝叶斯字典学习算法[J]. *通信学报*, 2019, 40(1): 43–50. doi: [10.11959/j.issn.1000-436x.2019015](https://doi.org/10.11959/j.issn.1000-436x.2019015).
- DONG Daoguang, RUI Guosheng, TIAN Wenbiao, *et al.* Nonparametric Bayesian dictionary learning algorithm based on structural similarity[J]. *Journal on Communications*, 2019, 40(1): 43–50. doi: [10.11959/j.issn.1000-436x.2019015](https://doi.org/10.11959/j.issn.1000-436x.2019015).
- [18] BISHOP C M. *Pattern Recognition and Machine Learning* (Information Science and Statistics)[M]. New York: Springer, 2006: 461–522.
- 董道广: 男, 1990年生, 博士, 研究方向为贝叶斯统计学习、压缩感知与大气波导.
- 芮国胜: 男, 1968年生, 博士, 教授, 博士生导师, 研究方向为现代通信理论及信号处理.
- 田文飏: 男, 1987年生, 博士, 副教授, 研究方向为大气波导与压缩感知.
- 张 洋: 男, 1983年生, 博士, 讲师, 研究方向为混沌通信技术.
- 刘 歌: 女, 1991年生, 博士, 研究方向为压缩感知与大气波导.
- 责任编辑: 陈 倩