

基于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, 康健, 刘歌

(海军航空大学信号与信息处理山东省重点实验室, 山东 烟台 264001)

摘 要: 相较于传统综合字典学习方法, 非参数贝叶斯方法具有显著优势, 但其对图像结构全局相似性和变异性的表示能力仍有较大提升空间。针对这个问题, 提出了一种基于结构相似性的非参数贝叶斯字典学习算法, 该算法基于图像结构的全局相似性对图像进行聚类处理, 并在图像的字典稀疏表示中引入块结构特性, 提升了字典的结构表示能力。实验表明, 所提算法在图像去噪和压缩感知方面的性能均优于目前主流的几种无监督字典学习算法。

关键词: 非参数贝叶斯; 字典学习; 结构相似性; 图像去噪; 压缩感知

中图分类号: TP301.6

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019015

Nonparametric Bayesian dictionary learning algorithm based on structural similarity

DONG Daoguang, RUI Guosheng, TIAN Wenbiao, KANG Jian, LIU Ge

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

Abstract: Though nonparametric Bayesian methods possesses significant superiority with respect to traditional comprehensive dictionary learning methods, there is room for improvement of this method as it needs more consideration over the structural similarity and variability of images. To solve this problem, a nonparametric Bayesian dictionary learning algorithm based on structural similarity was proposed. The algorithm improved the structural representing ability of dictionaries by clustering images according to their non-local structural similarity and introducing block structure into sparse representing of images. Denoising and compressed sensing experiments showed that the proposed algorithm performs better than several current popular unsupervised dictionary learning algorithms.

Key words: nonparametric Bayesian, dictionary learning, structural similarity, denoising, compressed sensing

1 引言

字典学习是图像和通信信号处理方面的重要内容, 是进行信号和图像数据分类^[1]、压缩^[2-6]、去噪^[7-11]、修复^[7,9-10,12-13]乃至超分辨^[14]的有力工具, 随着我国太空战略的深入推进, 字典学习在海洋信息感知领域, 尤其在卫星图像遥感、卫星气象遥感、卫星军事侦察等方面的应用价值愈发凸显。近年来, 非参数贝叶斯方法在字典学习中的应用引起人们的关注, 相对于以最优方向法 (MOD, method of optimal directions) 和 K 奇异值分解算法 (K-SVD,

K-singular value decomposition) 为代表的传统综合字典学习算法体现出了显著的优越性。该优越性体现在三方面: 1) 学习过程中能够自动推断出字典原子数目、信号稀疏度和正则化参数值, 从而无需对这些参数的取值进行预设, 避免了因预设不当而造成的人为影响; 2) 有严格的理论证明其算法的收敛性和解的最优性, 而 K-SVD 和 MOD 尚缺乏理论证明的支撑; 3) 其自身基于概率图模型进行建模, 结构清晰且具有开放性, 适于引入各种正则化概率先验约束。

Zhou 等^[11]将非参数贝叶斯用于图像去噪和压

收稿日期: 2018-05-31; 修回日期: 2018-12-31

基金项目: 国家自然科学基金资助项目 (No.41606117, No.41476089, No.61671016)

Foundation Item: The National Natural Science Foundation of China (No.41606117, No.41476089, No.61671016)

缩重构, 提出了 beta 过程因子分析(BPFA, beta process factor analysis)算法, 是贝叶斯因子分析方法在字典学习领域的一次成功尝试, 取得了较好的应用效果, 但该算法没有考虑到图像的全局结构相似性和变异性, 在字典结构特性方面仍有较大提升空间。后期围绕提升非参数贝叶斯字典学习的结构特性, 又陆续出现了狄氏过程-beta 过程因子分析(DP-BPFA, Dirichlet process-beta process factor analysis)^[7]和相依分层 beta 过程(dHBP, dependent hierarchical beta process)^[9-10]等算法, 这些算法将空间相关性引入了非参数贝叶斯建模, 使图像在字典学习下的稀疏表示能够体现出一定的邻近空间相似性, 显著提升了字典在去噪、修复、压缩感知等方面的应用性能, 但复杂度颇高, 运行时间较长。Sunil^[15-16]及 Li^[17-18]等认识到结构相似性和变异性对信号/图像稀疏表示的重要影响, 基于狄氏过程的聚类特性, 提出了适于多来源数据条件下的多字典学习算法, 较好地解决了图像稀疏表示对结构相似性的利用问题。但数据聚类的先验建模用到了 stick-breaking 模型, 因其非共轭的数学形式, 造成了后验推断上的非解析性。其次, 多字典学习适于数据来源非单一的情形, 对于单一来源数据内部结构的聚类效应, 其适用性仍值得进一步研究。在传统的结构字典下, 图像的稀疏表示具有块结构性, 这利于提高字典对信号结构特征的表达能, 且相似图像块稀疏表示的支撑集通常也相似, 利于进行多观测向量模型下的联合稀疏重构, 进而利于降低观测数目并提高重构性能。这为非参数贝叶斯字典学习提供了重要的启发, 即将字典学习建立在图像块相似性聚类的基础上, 并进一步将块结构特性引入稀疏表示作为字典学习的正则化约束条件。而非参数贝叶斯字典学习则因其对图像本身结构的自适应性, 继续保持其相对于传统结构字典的优势。

本文在上述问题和解决思路的启发下, 提出了一种基于结构相似性的非参数贝叶斯图像字典学习算法, 其主要创新有三方面: 1) 以图像块的相似性为测度, 通过 K-均值分类算法对图像块进行聚类处理, 将前述各非参数贝叶斯字典学习算法的空域结构特征从局部扩展到全局, 使所有具备相似结构的图像块共同参与本聚类图像块的稀疏表示; 2) 鉴于传统结构字典下多个相似信号联合稀疏重构时一般假设支撑集相同的局限性, 为使相似图像块的稀疏表示适应支撑集平滑变化的实际情况, 令同一

聚类内部的各图像块对字典原子的使用仅服从概率意义上的相同, 使得学习所获得的字典利于进行多信号的联合稀疏重构; 3) 将传统结构字典下信号稀疏表示普遍存在的块结构效应作为正则化约束条件引入非参数贝叶斯字典学习框架, 以提升字典对图像结构的表达能力。

2 贝叶斯建模

2.1 基于全局结构相似性的图像预处理

图像内部存在大量的相似冗余成分, 其在传统结构字典下的稀疏表示也具有相似性, 突出体现为支撑集的相似性。将全局相似性测度引入字典学习框架有利于提升字典对图像结构特征的表达能。本文采取 K-均值聚类方法进行图像块聚类。K-均值聚类是一种简单而又行之有效的无监督聚类方法, 在数据未分类亦无标签的条件下, 依预设的聚类数目随机选择相应的聚类中心, 通过计算各图像块到各聚类中心的相似性测度, 决定图像块的聚类归属, 随后将各聚类内部全部图像块的均值作为新的聚类中心, 再次计算各图像块到各聚类中心的相似性测度, 进行新一轮的聚类。迭代执行上述过程, 直至各聚类中心相对前次迭代的聚类中心无明显偏移或达到预定迭代数目上限时, 方结束迭代并输出最终聚类结果。其中最为关键的 2 个问题分别是聚类数目的确定和相似性测度的选择。聚类数目过多会造成聚类顽健性下降, 出现聚类碎片效应, 对字典学习造成不利的人为影响; 但也不宜过少, 因为过少会无法准确体现自然图像中丰富的结构信息。本文以图像块灰度值的欧式范数作为相似性测度, 并根据聚类分割中常用的贝叶斯信息准则(BIC, Bayesian information criterion)确定聚类数目。

2.2 非参数贝叶斯字典学习建模

设从图像中采集了 N 个 $P \times P$ 灰度块, 对其分别做列向量堆栈处理, 记作 $\mathbf{x}_i \in \mathbb{R}^{M \times 1}$, 其中 $M = P \times P$, $i = 1, 2, \dots, N$ 。此时经过前述聚类操作, 此时的数据集 $\{\mathbf{x}_i\}_{i=1}^N$ 是带有聚类标签的, 这将为后续的字典学习提供重要的结构参考。待学习的字典记作 $\mathbf{D} \in \mathbb{R}^{M \times K}$, 由 K 个与 \mathbf{x}_i 同维数的原子 $\{\mathbf{d}_i\}_{i=1}^K$ 构成。虽然不需要预置字典原子规模是非参数贝叶斯的一大优势, 但这与 K 值的选择并不矛盾, 因为非参数贝叶斯字典学习的核心思想是贝叶斯稀疏因子分析, 稀疏先验可以确保信号表示的稀疏性, 在

K 值足够大的条件下, 仍会存在相当数量的原子不参与对信号的表示, 很容易确定实际的原子规模。

本文算法的核心即基于聚类处理后的数据集 $\{\mathbf{x}_i\}_{i=1}^N$ 和下述非参数贝叶斯模型估计获得字典 \mathbf{D} 。

数据观测模型为

$$\mathbf{x}_i = \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i) + \boldsymbol{\varepsilon}_i, \quad i=1,2,\dots,N \quad (1)$$

$$\boldsymbol{\varepsilon}_i \sim \text{normal}(0, \alpha_0^{-1} \mathbf{I}_M) \quad (2)$$

其中, i 表示当前图像块在采集到的全部图像块中的位序; $\mathbf{s}_i \in \mathbb{R}^K$ 是稀疏表示的权重分量, 反映各原子参与表示信号的程度; $\mathbf{z}_i = \{0,1\}^K$ 是稀疏表示的模式分量, 反映各原子有无参与对信号的表示; 符号 \odot 是 Hadamard 乘积, 即逐元素对应乘积而不改变向量维数; $\boldsymbol{\varepsilon}_i$ 是均值为 0 的高斯白噪声, 其中 α_0 是观测噪声精度, 服从 Gamma 超先验分布, 如式(3)所示。

$$\alpha_0 \sim \text{Gamma}(\alpha_0 | a, b) \quad (3)$$

该分布与高斯分布共轭, a 和 b 分别表示形状参数和尺度参数, 为保证噪声幅值的稀疏性, 一般取 $a=10^{-6}$, $b=1$ 。

式(1)中, 当 \mathbf{z}_i 的某个元素 $z_{ik}=0$ 时, s_{ik} 便不再起作用, 因为原子 \mathbf{d}_k 并不参与对 \mathbf{x}_i 的表示。权重向量 \mathbf{s}_i 服从如式(4)所示的多元高斯先验分布。

$$\mathbf{s}_i \sim \text{normal}(\mathbf{s}_i | 0, \gamma_i^{-1} \mathbf{I}_K) \quad (4)$$

其中

$$\gamma \sim \text{Gamma}(\gamma | c, d) \quad (5)$$

出于防止过拟合的考虑, 令权重向量的全部元素使用相同的逆方差 γ 。类似地, Gamma 分布的参数一般取 $c=10^{-6}$, $d=1$ 。

与基于 BPFA 的算法相比, 本文所提算法将前述聚类过程获得的全局相似性结构特征引入到稀疏模式向量 \mathbf{z}_i 的先验分布中, 如式(6)和式(7)所示。

$$\mathbf{z}_i \sim \prod_{k=1}^K \text{Bernoulli}(z_{ik} | \pi_{jk}) \quad (6)$$

$$\pi_{jk} \sim \text{beta}(\pi_{jk} | \hat{e}, \hat{f}), \quad j=1,\dots,J \quad (7)$$

其中, π_{jk} 是聚类 j 中原子 \mathbf{d}_k 参与表示数据的概率, J 是聚类数目。

假设同一聚类内部的数据对字典原子的使用

概率是相同的, 允许这些数据不必共用同一支撑集, 而是仅在概率意义上具备支撑集的统一性, 从而适应了自然信号在空域和时域上普遍存在的支撑集缓变特性。需要指出的是, 假设当前数据所属聚类标签为 j , 该聚类内全部数据索引的集合记作 $\Omega(j)$, 则超参数 \hat{e} 和 \hat{f} 的取值决定于该聚类中全部数据稀疏模式向量相应元素 $\{z_{ik}\}_{i \in \Omega(j)}$ 的毗邻元素取值。引入符号 $\#\{\cdot\}$, 该符号代表花括号内集合的元素总数; $|\Omega(j)|$ 代表集合 $\Omega(j)$ 的元素总数, 并分别定义变量 Δ_1 、 Δ_2 和 Δ_3 , 如式(8)~式(9)所示, \hat{e} 和 \hat{f} 的取值规则如式(11)所示。

$$\Delta_1 = \#\{z_{i,k-1}=1, z_{i,k+1}=1, i \in \Omega(j)\} \quad (8)$$

$$\Delta_2 = \#\{z_{i,k-1}=0, z_{i,k+1}=0, i \in \Omega(j)\} \quad (9)$$

$$\Delta_3 = |\Omega(j)| - \Delta_1 - \Delta_2 \quad (10)$$

$$(\hat{e}, \hat{f}) = \begin{cases} (e^h, f^h), & \max\{\Delta_1, \Delta_2, \Delta_3\} = \Delta_1 \\ (e^l, f^l), & \max\{\Delta_1, \Delta_2, \Delta_3\} = \Delta_2 \\ (e^u, f^u), & \max\{\Delta_1, \Delta_2, \Delta_3\} = \Delta_3 \end{cases} \quad (11)$$

其中, $e^h = \frac{N-1}{N}$, $f^h = \frac{1}{N}$, $e^l = \frac{1}{N}$, $f^l = \frac{N-1}{N}$, $e^u = \frac{1}{2}$, $f^u = \frac{1}{2}$ 。

式(8)~式(11)的核心即将马尔可夫随机场的思想引入字典学习框架中。当邻域稀疏模式均为 1 时, z_{ik} 取值为 1 的概率更大, 令 $e^h \gg f^h$ 有利于生成数值较大的 π_{jk} , 从而倾向于使 z_{ik} 以较大概率取值为 1; 当邻域稀疏模式均为 0 时, z_{ik} 取值为 0 的概率更大, 令 $e^l \ll f^l$ 有利于生成数值较小的 π_{jk} , 从而倾向于使 z_{ik} 以较大概率取值为 0; 当邻域稀疏模式为 1 和 0 共存时, z_{ik} 具备取值的不确定性, 此时令 $e^u = f^u$ 可以保证 z_{ik} 取值概率的均衡性。

然而, 有别于各数据向量具有不同原子使用概率的情形, 也有别于所有数据向量共用同一原子使用概率的情形, 全局结构相似性聚类后的数据在同一聚类内共用相同的原子使用概率, 这是对 z_{ik} 取值的折中处理, 也是结构聚类思想的本质要求, 为此, 需统计聚类内部各数据向量的邻域稀疏模式, 以数量最多的邻域稀疏模式决定 \hat{e} 和 \hat{f} 的取值, 如式(11)所示。字典原子 $\{\mathbf{d}_k\}_{k=1}^K$ 服从如式(12)

所示的高斯先验分布。

$$\mathbf{d}_k \sim \text{normal}(\mathbf{d}_k | 0, M^{-1} \mathbf{I}_M), k = 1, 2, \dots, K \quad (12)$$

基于上述理论，建立贝叶斯建模的概率图模型，如图1所示。

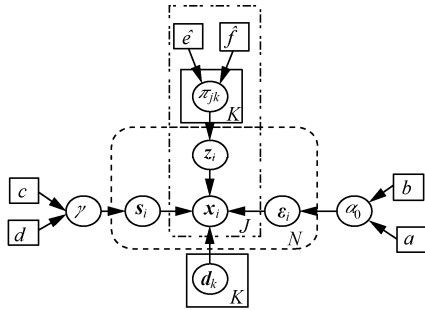


图1 贝叶斯建模的图模型

3 贝叶斯推断

上述模型中，除聚类处理后的观测数据集 $\{\mathbf{x}_i\}_{i=1}^K$ 和超参数已知外，其余变量和模型参数均是未知量，其中包括隐变量集 $\Theta = \{\{s_i\}_{i=1}^N, \{z_i\}_{i=1}^N, \{\pi_{jk}, j=1, \dots, J\}_{k=1}^K, \{\mathbf{d}_k\}_{k=1}^K\}$ 和模型参数集 $H = \{\gamma, \alpha_0\}$ ，可借助 MCMC Gibbs 采样器^[19]对其进行推断估计，这是因为 MCMC 方法适用性广泛，且相对简单易行，作为一种典型的贝叶斯参数估计方法，它不但给出参数的后验均值估计，还以概率分布的形式给出该估计的置信水平。所以本文设计 Gibbs 采样器进行非参数贝叶斯字典学习的变量推断，下述给出待推断变量的采样公式及相关推导。

3.1 原子 \mathbf{d}_k 的 Gibbs 采样

结合式(1)观测数据模型和式(12)原子先验分布，可得 \mathbf{d}_k 满条件分布，如式(13)所示。

$$p(\mathbf{d}_k | -) \propto \prod_{i=1}^N \text{normal}(\mathbf{x}_i | \mathbf{D}(s_i \odot \mathbf{z}_i), \alpha_0^{-1} \mathbf{I}_M) \cdot \text{normal}(\mathbf{d}_k | 0, M^{-1} \mathbf{I}_M) \quad (13)$$

其中

$$\text{normal}(\mathbf{x}_i | \mathbf{D}(s_i \odot \mathbf{z}_i), \alpha_0^{-1} \mathbf{I}_M) = (2\pi\alpha_0^{-1})^{\frac{M}{2}} e^{-\frac{\alpha_0}{2} \left\| \mathbf{x}_i - \sum_{l=1}^K s_{il} z_{il} \mathbf{d}_l \right\|_2^2} \quad (14)$$

记 $\tilde{\mathbf{x}}_{i,-k} = \mathbf{x}_i - \sum_{l \neq k} s_{il} z_{il} \mathbf{d}_l$ ，则式(14)等号右侧的指

数项内的 $\left\| \mathbf{x}_i - \sum_{l=1}^K s_{il} z_{il} \mathbf{d}_l \right\|_2^2$ 即变换为 $\left\| \tilde{\mathbf{x}}_{i,-k} - s_{ik} z_{ik} \mathbf{d}_k \right\|_2^2$ ，

将其代入(13)式并整理，可得

$$p(\mathbf{d}_k | -) \propto \text{normal}(\mathbf{d}_k | \boldsymbol{\mu}_{\mathbf{d}_k}, \boldsymbol{\Sigma}_{\mathbf{d}_k}) \quad (15)$$

$$\boldsymbol{\mu}_{\mathbf{d}_k} = \alpha_0 \left(M + \alpha_0 \sum_{i=1}^N s_{ik}^2 z_{ik} \right)^{-1} \sum_{i=1}^N s_{ik} z_{ik} \tilde{\mathbf{x}}_{i,-k} \quad (16)$$

$$\boldsymbol{\Sigma}_{\mathbf{d}_k} = \left(M + \alpha_0 \sum_{i=1}^N s_{ik}^2 z_{ik} \right)^{-1} \mathbf{I}_M \quad (17)$$

3.2 权重 s_i 的 Gibbs 采样

结合式(1)观测数据模型和式(4)权重先验分布，可得 s_{ik} 的满条件分布，如式(18)所示。

$$p(s_{ik} | -) \propto \text{normal}(\mathbf{x}_i | \mathbf{D}(s_i \odot \mathbf{z}_i), \alpha_0^{-1} \mathbf{I}_M) \cdot \text{normal}(s_{ik} | 0, \gamma^{-1}) \quad (18)$$

整理可得

$$p(s_{ik} | -) \propto \text{normal}(s_{ik} | \mu_{s_{ik}}, \Sigma_{s_{ik}}) \quad (19)$$

$$\mu_{s_{ik}} = \alpha_0 z_{ik} \left(\gamma + \alpha_0 z_{ik} \mathbf{d}_k^T \mathbf{d}_k \right)^{-1} \tilde{\mathbf{x}}_{i,-k}^T \mathbf{d}_k \quad (20)$$

$$\Sigma_{s_{ik}} = \left(\gamma + \alpha_0 z_{ik} \mathbf{d}_k^T \mathbf{d}_k \right)^{-1} \quad (21)$$

据此对 s_{ik} ($k=1, 2, \dots, K$) 依次采样即实现对权重向量 \mathbf{s}_i 的 Gibbs 采样。

3.3 稀疏模式 z_i 的 Gibbs 采样

结合式(1)观测数据模型与式(6)稀疏模式先验分布，可得 z_{ik} 的满条件后验分布，如式(22)所示。

$$p(z_{ik} | -) \propto \text{normal}(\mathbf{x}_i | \mathbf{D}(z_i \odot \mathbf{s}_i), \alpha_0^{-1} \mathbf{I}_M) \cdot \text{Bernoulli}(z_{ik} | \pi_{jk}) \quad (22)$$

当 $z_{ik} = 1$ 时，记

$$p(z_{ik} | -) \propto \pi_{jk} e^{\alpha_0 s_{ik} \tilde{\mathbf{x}}_{i,-k}^T \mathbf{d}_k - \frac{\alpha_0}{2} s_{ik}^2 \mathbf{d}_k^T \mathbf{d}_k} = p_1 \quad (23)$$

当 $z_{ik} = 0$ 时，记

$$p(z_{ik} | -) \propto 1 - \pi_{jk} = p_0 \quad (24)$$

整理可得 z_{ik} 的满条件分布为 Bernoulli 分布，即

$$p(z_{ik} | -) \propto \text{Bernoulli}(z_{ik} | \tilde{\pi}_{jk}) \quad (25)$$

$$\tilde{\pi}_{jk} = \frac{p_1}{p_1 + p_0} \quad (26)$$

据此对 z_{ik} ($i=1, \dots, N; k=1, \dots, K$) 依次采样实现对稀疏模式向量 \mathbf{z} 的 Gibbs 采样, π_{jk} 的更新是本聚类意义上的更新, 当 $z_{i+1,k}$ 属于另一个聚类 j' 时, 其相应 $\pi_{j'k}$ 的更新是对前一个 $\pi_{j'k}$ 取值的更新, 与 π_{jk} 无关。

3.4 原子使用概率 π 的 Gibbs 采样

结合式(6)和式(7), 可得 π_{jk} 的满条件分布, 如式(27)所示。

$$p(\pi_{jk} | -) \propto \prod_{i \in \Omega(j)} \text{Bernoulli}(z_{ik} | \pi_{jk}) \text{beta}(\pi_{jk} | \hat{e}, \hat{f}) \quad (27)$$

其中, $\Omega(j)$ 是聚类 j 内全部数据的索引集合, \hat{e} 和 \hat{f} 的取值参考式(8)~式(11)。式(27)经整理可得仍是关于 π_{jk} 的 beta 分布, 即

$$p(\pi_{jk} | -) \propto \text{Beta}(\pi_{jk} | \tilde{e}, \tilde{f}) \quad (28)$$

$$\tilde{e} = \sum_{i \in \Omega(j)} z_{ik} + \hat{e} \quad (29)$$

$$\tilde{f} = |\Omega(j)| - \sum_{i \in \Omega(j)} z_{ik} + \hat{f} \quad (30)$$

因此, 对 π_{jk} ($j=1, \dots, J; k=1, \dots, K$) 依次采样实现对原子使用概率向量 π 的 Gibbs 采样。同样地, 式(29)和式(30)对 \tilde{e} 和 \tilde{f} 的更新也是局限在本聚类意义上的, 对于另一聚类原子使用概率, 其超参数的更新则与本聚类无关。

3.5 超参数 α_0 和 γ 的 Gibbs 采样

结合式(1)和式(3), 可得超参数 α_0 的满条件分布, 如式(31)所示。

$$p(\alpha_0 | -) \propto \prod_{i=1}^N \text{normal}(\mathbf{x}_i | \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i), \alpha_0^{-1} \mathbf{I}_M) \text{Gamma}(\alpha_0 | a, b) \quad (31)$$

整理可得

$$p(\alpha_0 | -) \propto \text{Gamma}(\alpha_0 | \tilde{a}, \tilde{b}) \quad (32)$$

$$\tilde{a} = a + \frac{MN}{2} \quad (33)$$

$$\tilde{b} = b + \frac{1}{2} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i)\|_2^2 \quad (34)$$

结合式(4)和式(5), 可得超参数 γ 的满条件分布, 如式(35)所示。

$$p(\gamma | -) \propto \prod_{i=1}^N \text{normal}(\mathbf{s}_i | 0, \gamma^{-1} \mathbf{I}_K) \cdot \text{Gamma}(\gamma | c, d) \quad (35)$$

整理可得

$$p(\gamma | -) \propto \text{Gamma}(\gamma | \tilde{c}, \tilde{d}) \quad (36)$$

$$\tilde{c} = c + \frac{KN}{2} \quad (37)$$

$$\tilde{d} = d + \frac{1}{2} \sum_{i=1}^N \mathbf{s}_i^T \mathbf{s}_i \quad (38)$$

根据式(32)~式(34)可对超参数 α_0 进行 Gibbs 采样, 根据式(36)~式(38)即可实现对超参数 γ 的 Gibbs 采样。

迭代执行本节的 Gibbs 采样流程, 直至达到预定数目的迭代上限, 即结束迭代并输出各隐变量和模型参数的当前采样值作为推断结果, 完成字典学习, 本文与传统 BPFA 保持一致, 将迭代次数上限设置为 150, 为免混淆, 本文算法称为结构相似性 beta 过程因子分析 (SSIM-BPFA, structure similarity-beta process factor analysis) 算法, 具体流程见算法 1。

算法 1 SSIM-BPFA 字典学习算法

输入 图像块数据集 $\{\mathbf{x}_i\}_{i=1}^N$ 。

初始化 超参数 $a=10^{-6}$, $b=1$, $c=10^{-6}$, $d=1$, 当图像大小超过 300×300 时令字典原子规模 $K=512$, 否则令 $K=256$ 。

步骤 1 利用 K-均值算法结合传统的 BIC 准则对 $\{\mathbf{x}_i\}_{i=1}^N$ 进行聚类预处理, 确定聚类数目 J , 并获得每个数据点的聚类标签 j , $j=1, \dots, J$ 。

步骤 2 根据式(1)~式(12)进行贝叶斯概率建模, 建立隐变量和模型参数集 $\Theta = \{\{\mathbf{d}_k\}_{k=1}^K, \{\mathbf{s}_i\}_{i=1}^N, \{\mathbf{z}_i\}_{i=1}^N, \{\pi_{jk}\}_{j=1, k=1}^{J, K}, \alpha_0, \gamma\}$ 。

步骤 3 设置迭代总数为 150, 为当前迭代索引赋值 $t=1$ 。

步骤 4 按照式(15)~式(17)对字典原子 $\{\mathbf{d}_k\}_{k=1}^K$ 进行 Gibbs 采样。

步骤 5 按照式(19)~式(21)对权重 $\{\mathbf{s}_i\}_{i=1}^N$ 进行 Gibbs 采样。

步骤 6 按照式(23)~式(26)对稀疏模式 $\{\mathbf{z}_i\}_{i=1}^N$ 进行 Gibbs 采样。

步骤 7 按照式(28)~式(30), 并参考式(8)~

式(11)对原子使用概率 $\{\pi_{jk}\}_{j=1,k=1}^{J,K}$ 进行 Gibbs 采样。

步骤 8 分别按照式(32)~式(34)和式(36)~式(38)对超参数 α_0 和 γ 进行后验 Gibbs 采样。

步骤 9 若 $t=150$ ，则将 Θ 内各元素的当前采样值作为估计值；若 $t \neq 150$ ，则令 $t=t+1$ ，并返回步骤 4 执行迭代。

输出 当前估计的字典 $D = \{d_k\}_{k=1}^K$ ，权重 $\{s_i\}_{i=1}^N$ 和稀疏模式 $\{z_i\}_{i=1}^N$ 。

4 算法分析与数值实验

本文围绕图像去噪和压缩感知这 2 个方面的性能对所提的 SSIM-BPFA 算法的有效性展开评估，以图像处理领域中的 Barbara、Lena、Boath、House 和 Peppers 图像的灰度值作为实验对象，使用 Matlab 编程进行数值实验。

4.1 去噪实验及性能评估

以分辨率 512 像素 \times 512 像素的图像 Barbara 为例，采用滑窗分块方式将其分割为 255 025 个具有重叠的 8 像素 \times 8 像素分辨率的图像块，并对全部图像块的灰度值进行列堆栈处理，构成算法执行所使用的数据集。按照 SSIM-BPFA 算法的流程进行字典学习，并将初始字典原子数目设置为 512 个，训练所得字典如图 2 所示。

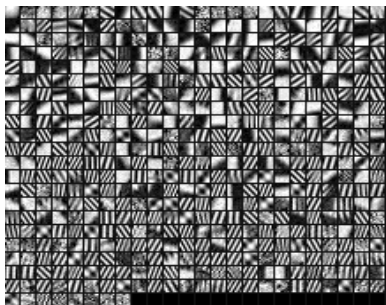


图 2 SSIM-BPFA 算法训练所得图像字典

设噪声标准差 $\sigma = 25$ ，采用 SSIM-BPFA 算法

对 Barbara 进行去噪处理，结果如图 3 所示。可见相比于图 3(b)所示的含噪图像，图 3(c)所示的去噪图像的峰值信噪比 (SNR, signal noise ratio) 显著提高，图像视觉效果也得到了明显的改善，而且运行时间较 BPFA 算法并未有明显的增加。



图 3 SSIM-BPFA 算法去噪实验结果

为进一步验证 SSIM-BPFA 算法的性能，除 Barbara 外又增加了 Lena、Boat、House、Peppers 等作为实验图像，并引入 MOD、K-SVD、传统的离散余弦傅里叶变换 (DCT, discrete cosine transformation) 等传统的字典学习算法与本文的 SSIM-BPFA 算法进行比较。实验过程中，分别采用噪声的标准差为 5、10、15 和 25 这 4 种参数进行实验。实验结果显示，噪声标准差等于 5、10、15 时 7 种算法的去噪结果与表 1 相似，只是在噪声标准差较大时对比更为显著。因此，本文以 $\sigma = 25$ 为例，对比了 SSIM-BPFA 和其他 6 种算法对实验图像去噪处理后的峰值信噪比结果，如表 1 所示，表中的实验结果是进行 50 次实验后，实验结果的平均值。

由表 1 可知，SSIM-BPFA 的去噪峰值信噪比明显高于其他 6 种算法，其中 BPFA、DP-BPFA 和 dHBP 这 3 种算法的去噪性能相近，这与 Zhou 的研究结论是一致的^[7]。

4.2 压缩感知实验及性能评估

利用前述去噪过程中 7 种不同算法分别学习获得的字典，结合相应的压缩感知算法，对经观测矩

表 1 不同字典学习算法的图像去噪峰值信噪比

实验图像	SSIM-BPFA 算法/dB	BPFA 算法/dB	DP-BPFA 算法/dB	dHBP 算法/dB	KSVD 算法/dB	MOD 算法/dB	DCT 算法/dB
Barbara	29.924 0	28.747 0	28.783 0	28.789 3	28.731 4	28.599 5	24.172 4
Lena	30.261 7	29.934 0	29.942 5	29.967 1	29.892 2	29.639 1	25.832 1
Boat	29.713 4	28.592 1	28.620 1	28.632 0	28.485 2	28.201 4	24.480 1
House	30.350 7	29.987 3	29.987 5	29.990 3	29.896 3	29.709 6	24.922 0
Peppers	29.662 9	28.581 0	28.587 0	29.012 2	28.552 8	28.247 0	25.938 1

阵降维投影后的图像数据进行重构。Zhou^[7]的研究已经证实了 BPFA 算法相对于其他字典学习算法在压缩感知应用中的优势，所以限于篇幅，本节实验着重对 SSIM-BPFA 算法所得字典与 BPFA、DP-BPFA 和 dHBP 这 3 种算法所得字典进行压缩感知性能对比。为便于比较分析，仍以 Barbara 作为实验对象，出于克服块结构效应的重构需求，实验数据的生成方式仍采用 4.1 节的滑窗法。由于前述字典学习过程中，图像块在字典表示下的稀疏度普遍不超过 4，故出于提高重构成功概率的目的，本节实验选择的投影观测矩阵行数以 8 为下限，设置了行数分别为 8、10、12、14、16、18 等共 6 种观测数目作为实验条件，所用投影观测矩阵的元素独立同分布地采样自标准正态分布。

图 4 所示即观测数目为 8 时，压缩感知经典的正交匹配追踪(OMP, orthogonal matching pursuit)算法分别在 SSIM-BPFA 算法、BPFA 算法、DP-BPFA 算法和 dHBP 算法所得字典下对 Barbara 的重构结果及相应的峰值信噪比。由图可见，图 4(d)在视觉效果和峰值信噪比方面均显著优于图 4(a)~图 4(c)。在 Lena 等其他 4 幅实验图像的压缩感知实验中，SSIM-BPFA 算法所得字典也体现出了同样的优越性。



图 4 Barbara 重构结果及峰值信噪比

为了进一步验证不同投影观测数目条件下 SSIM-BPFA 所得字典的压缩感知性能，图 5 给出了观测数目分别为 8、10、12、14、16、18 等共 6 种条件下，在 SSIM-BPFA 算法、dHBP 算法、DP-BPFA

算法、BPFA 算法和 OMP 算法所得字典表示下分别通过 OMP 算法对 Barbara 的重构相对误差，数据为 50 次实验结果的平均值。

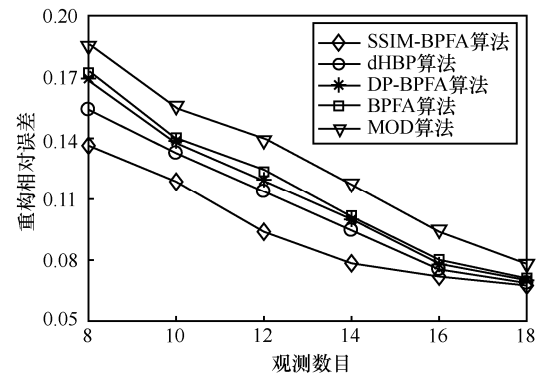


图 5 不同算法所得字典下的重构相对误差

从图 5 可以看出，随着观测数目的增加，不同算法字典下的重构相对误差均呈下降趋势，但总体重构相对误差由低到高分别是 SSIM-BPFA 算法、dHBP 算法、DP-BPFA 算法、BPFA 算法和 MOD 算法所得字典，其中，前四种算法重构相对误差均显著低于 MOD 算法字典的重构相对误差，且 SSIM-BPFA 字典的重构相对误差显著低于其他字典的重构相对误差，这从另一个侧面表明 SSIM-BPFA 算法学习所得字典在压缩感知中的应用性能优于其他几种算法所得字典。其原因在于 SSIM-BPFA 算法基于全局结构相似性对图像进行了聚类处理，并引入了稀疏表示的块结构特性，使得字典对图像结构特征的表示性能得以改善，图像表示的稀疏度更小，从而提高了低观测数目条件下的重构成功率。尽管 Gibbs 采样推断方法的理论收敛性不易证明，但关于该推断方法收敛效果的可靠性和有效性均已在前人研究成果^[9-12]中得到了实验验证。

5 结束语

本文提出一种基于结构相似性的非参数贝叶斯字典学习算法，将马尔可夫随机场思想引入到图像稀疏表示的块结构建模中，使字典具备了传统结构字典稀疏表示的块结构特性。实验表明，该算法所得字典的去噪性能和压缩感知性能优于 BPFA 算法、DP-BPFA 算法和 dHBP 算法这 3 种非参数贝叶斯算法字典。本文以后将在两方面进行下一步研究：1) Gibbs 采样推断计算量相对较大，这是 BPFA 算法同样面临的问题，且采样推断收敛性能的可靠性和有效性不易于理论证明，后续将研究效率更高

的推断算法, 并着重加强理论证明; 2) 图像结构相似性聚类过程采用 K-means 算法执行, 虽简单易行复杂度低, 但需要进行聚类数目的验证性优化, 限制了算法的自适应性和顽健性, 后续应着重提升无监督聚类的自适应性, 以期在可观的复杂度之外, 进一步改善字典的应用性能。

参考文献:

- [1] FEDOROV I, RAO B D, NGUYEN T Q. Multimodal sparse Bayesian dictionary learning applied to multimodal data classification[C]// IEEE International Conference on Acoustics.2017: 2237-2241.
- [2] HUANG Y, PAISLEY J, LIN Q, et al. Bayesian nonparametric dictionary learning for compressed sensing MRI[J]. IEEE Transactions on Image Processing, 2014, 23(12): 5007-5019.
- [3] LI S, TAO X, LU J. Variational Bayesian inference for nonparametric signal compressive sensing on structured manifolds[C]//IEEE International Conference on Communications.2017: 1-6.
- [4] LI Z, HUANG H, MISRA S. Compressed sensing via dictionary learning and approximate message passing for multimedia Internet of things[J]. IEEE Internet of Things Journal, 2017, 4(2): 505-512.
- [5] HUANG Y, PAISLEY J, LIN Q, et al. Bayesian nonparametric dictionary learning for compressed sensing MRI[J]. IEEE Transactions on Image Processing, 2014, 23(12): 5007-5019.
- [6] HONG P D, CHAINAIS P. A Bayesian non-parametric approach to learn dictionaries with adapted numbers of atoms[C]//IEEE International Workshop on Machine Learning for Signal Processing.2015: 1-6.
- [7] ZHOU M, CHEN H, 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.
- [8] LIU Z, YU L, ZHANG M, et al. Nonlocal structured nonparametric Bayesian dictionary learning for image denoising[C]//IEEE International Conference on Signal Processing.2017: 144-148.
- [9] ZHOU M, YANG H, SAPIRO G, et al. Dependent hierarchical beta process for image interpolation and denoising[J]. Aistats, 2012, 15: 883-891.
- [10] PAISLEY J, ZHOU M, SAPIRO G, et al. Nonparametric image interpolation and dictionary learning using spatially-dependent dirichlet and beta process priors[C]// IEEE International Conference on Image Processing.2010: 1869-1872.
- [11] ZHOU M, CHEN H, PAISLEY J, et al. Non-parametric Bayesian dictionary learning for sparse image representations[C]//International Conference on Neural Information Processing Systems. 2009, 21(1): 2295-2303.
- [12] ZHOU G, ZHU D, WANG K, et al. Wavelet image inpainting based on dictionary learning with a Beta process[C]//Proceedings of 2010 Conference on Dependable Computing.2010: 83-88.
- [13] ZHU L, HUANG Z, LIU Y, et al. The nonparametric Bayesian dictionary learning based interpolation method for WSN missing data[J]. AEU-International Journal of Electronics and Communications, 2017(79): 267-274.
- [14] HE L, QI H, ZARETZKI R. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution[J]. Computer Vision & Pattern Recognition, 2013, 9(4): 345-352.
- [15] GUPTA S K, PHUNG D, ADAMS B, et al. A Bayesian framework for learning shared and individual subspaces from multiple data sources[C]//Pacific-Asia Conference on Advances in Knowledge Discovery & Data Mining. 2011, 6634(1): 136-147.
- [16] GUPTA S K, PHUANG D, VENKATESH S. A Bayesian nonparametric joint factor model for learning shared and individual subspaces from multiple data sources[C]//Proc of Siam Int Conference on Data Mining. 2013:200-212.
- [17] LI S, TAO X, LU J. Variational inference for nonparametric subspace dictionary learning with hierarchical Beta process[C]//IEEE International Conference on Acoustics .2017: 2691-2695.
- [18] LI S, TAO X, DONG L. A nonparametric Bayesian approach to joint multiple dictionary learning with separate image sources[J]. Signal & Information Processing , 2016 :1155-1159.
- [19] CHASPARI T, TSILIFIS P. MCMC inference of parametric dictionaries for sparse Bayesian approximations[J]. IEEE Transactions on Signal Processing, 2016, 64 (12): 3077-3092.

[作者简介]



董道广(1990-), 男, 山东济南人, 海军航空大学博士生, 主要研究方向为 Bayesian 统计学习、压缩感知和蒸发波导反演。



芮国胜(1968-), 男, 山东烟台人, 博士, 海军航空大学教授、博士生导师, 主要研究方向为混沌通信系统及现代滤波理论。



田文颺(1987-), 男, 江西南昌人, 博士, 海军航空大学副教授, 主要研究方向为压缩感知、蒸发波导反演。

康健(1971-), 女, 黑龙江哈尔滨人, 博士, 海军航空大学副教授, 主要研究方向为信号处理及现代滤波理论。

刘歌(1991-), 女, 山东威海人, 海军航空大学博士生, 主要研究方向为压缩感知、蒸发波导反演。