GlobalSIP 2015

A Nonparametric Bayesian Approach to Joint Multiple Dictionary Learning with Separate Image Sources

Shaoyang Li, Xiaoming Tao, Linhao Dong, and Jianhua Lu

Department of Electronic Engineering

Tsinghua University, Beijing, China

December 16th, 2015

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- 4. Experimental Results
- 5. Conclusions

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- 4. Experimental Results
- 5. Conclusions

Research Background

- **Starting Point —— Dictionary Learning**

✓ Dictionary learning provides a framework of sparse representations for **high-dimensional signals** (*e.g.*, images)

✓ By seeking for the closest matching dictionary, the images of interest can be represented as the superposition of **small subsets of dictionary atoms**

Dictionary Dimension

→ **Fixed** number of dictionary atoms

→ **Infer** the required number of atoms

nonparametric Bayesian methods

Research Background

- For the applications where image samples are extracted from **multiple sources or categories**

e.g. image patches representing **textures of different animals**




The image patches from **different animal categories** may belong to different low-dimensional **subspaces or manifolds**

Multiple Dictionary Learning

Research Background

- Existing Bayesian approaches still represent image samples with **a unified dictionary**

$$\boxed{x = \mathbf{D}w + \epsilon} \Rightarrow \begin{array}{l} x = \mathbf{D}^* w + \epsilon \\ \mathbf{D}^* \in \{\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \ldots\} \end{array}$$

A set of dictionaries for image samples should be optimally learned

- **Problem ①**

  for the number of dictionaries

  - Some works have introduced a **clustering setting** into dictionary learning
  - Implementing such a strategy have to **determine a fixed number** for dictionaries in advance

Research Background

Depict correlation of dictionaries → Proper Bayesian prior → Hierarchical Beta process

- **Problem ②**
  - Due to non-conjugacy limitations, the inference for Bayesian models with HBP is often intractable
  - Existing inference approaches for HBP may require assumptions or approximations

---

Contributions of this paper

- Dictionary Number → Dirichlet process
- Dictionary Correlations → Inference engine

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- 4. Experimental Results
- 5. Conclusions

Problem Formulation

➤ **Discover a sparse representation spanned by the atoms :**

$$\boxed{\boldsymbol{x}_i = \mathbf{D}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i} \Rightarrow \boxed{\boldsymbol{x}_i = \mathbf{D}_{c(i)}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i}$$

$\{\boldsymbol{x}_i\}_{i=1}^N \in \mathbb{R}^P, \mathbf{D} \in \mathbb{R}^{P\times K}, \boldsymbol{w}_i \in \mathbb{R}^K$ $\quad c(i) \in \{1, 2, 3, \dots\}$

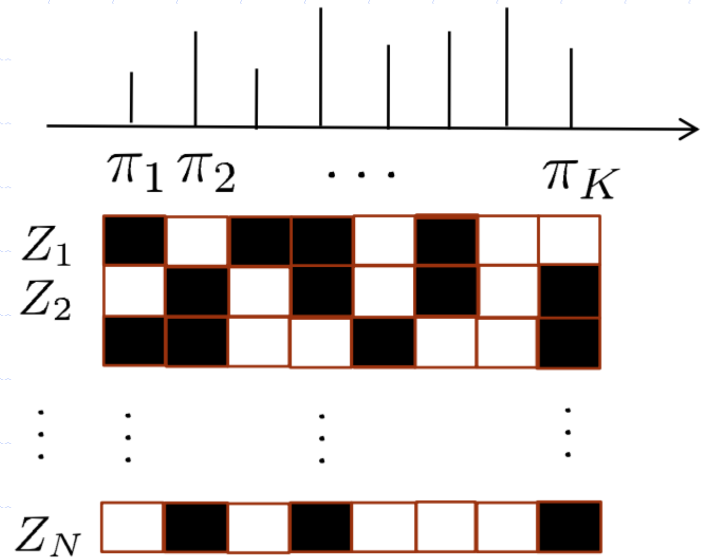
Basic model in the framework of Bayesian learning：

$$\boldsymbol{x}_i \sim \mathcal{N}(\mathbf{D}_{c(i)}\boldsymbol{w}_i, \alpha_{c(i)}^{-1}\mathbf{I}_P),$$

$$\mathbf{D}_{c(i)} \sim \prod_{k=1}^{K} \mathcal{N}(\mathbf{0}, \frac{1}{P}\mathbf{I}_P),$$

$$\boldsymbol{w}_i = \boldsymbol{z}_{c(i)} \odot \boldsymbol{s}_i, \;\; \boldsymbol{s}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K),$$

$$\boldsymbol{z}_i|\boldsymbol{\pi}_c \sim \prod_{k=1}^{K} \mathrm{Bernoulli}(\pi_{ck}), \;\; \forall\ i : c(i) = c,$$



Problem Formulation

- **Bayesian Prior ① : Dirichlet process (DP)**

$$c(i) \sim G = \sum_{c=1}^{\infty} \xi_c \delta_{G_c}, \quad G \sim \mathcal{DP}(\eta, \{G_c\}_{c=1}^{\infty})$$

→ **Unbounded dictionary number**

$$\boldsymbol{\xi} \sim \mathrm{GEM}(\eta), \ \eta \sim \mathrm{Gamma}(a, b)$$

$$\xi_c = \rho_c \prod_{l=1}^{c-1} (1 - \rho_l), \ \rho_c \sim \mathrm{Beta}(1, \eta)$$

→ **Stick-breaking construction**

- **Bayesian Prior ② : Hierarchical Beta process (HBP)**

$$G_c = \sum_{k=1}^{\infty} \pi_{ck} \delta_{\phi_k} \overset{\text{iid}}{\sim} \mathcal{BP}(\gamma_c, H),$$

$$\pi_{ck} | \gamma_c, \upsilon \sim \mathrm{Beta}(\gamma_c \upsilon_k, \gamma_c \bar{\upsilon}_k),$$

$$H = \sum_{k=1}^{\infty} \upsilon_k \delta_{\phi_k} \sim \mathcal{BP}(\lambda, H_0), \ \phi_k \overset{\text{iid}}{\sim} H_0,$$

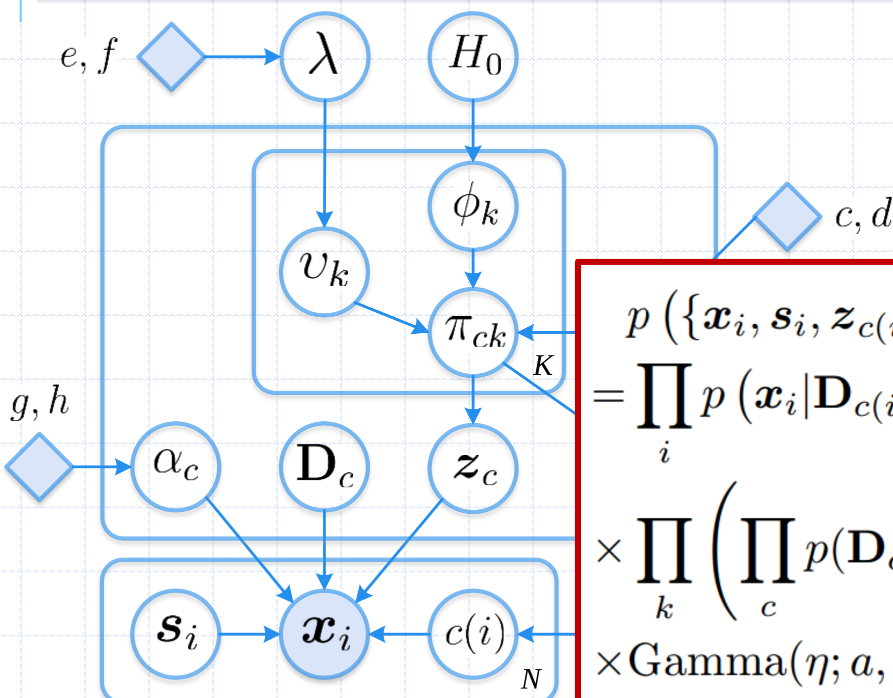
→ **Dictionary correlation priors**

Problem Formulation

- **Priors for hyper-parameters:**

$$v_k = \prod_{j=1}^{k} \beta_j, \;\; \beta_j \sim \mathrm{Beta}(\lambda, 1)$$ **Indian buffet process**

$$\gamma_c \sim \mathrm{Gamma}(c, d), \;\; \lambda \sim \mathrm{Gamma}(e, f), \;\; \alpha_c \sim \mathrm{Gamma}(g, h)$$



Factorizing joint distribution via probabilistic graphical model

$$
\begin{aligned}
& p\left(\{\boldsymbol{x}_i, \boldsymbol{s}_i, \boldsymbol{z}_{c(i)}, c(i)\}_{i=1}^{N}, \{\mathbf{D}_c, \boldsymbol{\pi}_c, \alpha_c\}_{i=1}^{C}, \boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \eta, \lambda\right) \\
&= \prod_i p\left(\boldsymbol{x}_i | \mathbf{D}_{c(i)}, \boldsymbol{z}_{c(i)}, \boldsymbol{s}_i, \alpha_c\right) p(\boldsymbol{s}_i) p\left(c(i) | \boldsymbol{\rho}\right) \times \prod_c p(\rho_c | \eta) \\
&\times \prod_k \left(\prod_c p(\mathbf{D}_{ck}) p(z_{ck} | \pi_{ck}) p(\pi_{ck} | \gamma_c, v_k) \times p(\alpha_c) \right) p(v_k | \lambda) \\
&\times \mathrm{Gamma}(\eta; a, b) \mathrm{Gamma}(\boldsymbol{\gamma}; c, d) \mathrm{Gamma}(\lambda; e, f),
\end{aligned}
$$

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- 4. Experimental Results
- 5. Conclusions

Learning Strategy via Sampling

Non-conjugate prior Bayesian Model → No closed-form solutions

Simplified Model → Assumptions or approximations

Markov chain Monte Carlo (MCMC) Sampling →

Samling for the DP

collapsed Gibbs sampling

+

Sampling for the HBP

auxiliary-variable-based slice sampling

Learning Strategy via Sampling

Sampling for the DP

① **Posterior for *c(i)***

$$
\begin{aligned}
&p(c(i) = c \,|\, \xi_c, \mathbf{D}_c, \boldsymbol{z}_c, \alpha_c) \\
&\propto \xi_c \int \mathcal{N}(\boldsymbol{x}_i; \mathbf{D}_{c(i)} \mathrm{diag}(\boldsymbol{z}_{c(i)}) \boldsymbol{s}_i, \alpha_c^{-1} \mathbf{I}_P) \mathcal{N}(\boldsymbol{s}_i; \mathbf{0}, \mathbf{I}_K) \mathrm{d}\boldsymbol{s}_i \\
&\propto \xi_c \exp\left\{ \boldsymbol{x}_i^{\mathsf{T}} (\alpha_c^{-1} \mathbf{I}_P + \widetilde{\mathbf{D}}_c^{\mathsf{T}} \widetilde{\mathbf{D}}_c)^{-1} \boldsymbol{x}_i \right\} \propto \xi_c \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c),
\end{aligned}
$$

② **Posterior for hyper-parameters in DP**

$$
\begin{aligned}
p(\eta|-) &\propto \prod_{c=1}^{C-1} \left(\tfrac{\Gamma(\eta+1)}{\Gamma(1)\Gamma(\eta)} (1-\rho_c)^{\eta-1} \right) \eta^{a-1} e^{-b\eta} \\
&\propto \mathrm{Gamma}\left(\eta; a + C - 1, b - \textstyle\sum_{c=1}^{C-1} (1-\rho_c) \right), \\
p(\rho_c|-) &\propto \prod_{i:c(i)=c} \rho_c \prod_{i:c(i)>c} (1-\rho_c)(1-\rho_c)^{\eta-1} \\
&\propto \mathrm{Beta}\left(\rho_c; 1 + \textstyle\sum_{i=1}^{N} \mathbb{I}(c(i) = c), \eta + \sum_{i=1}^{N} \mathbb{I}(c(i) > c) \right),
\end{aligned}
$$

Learning Strategy via Sampling

Sampling for the HBP

① Integrating out $\{\boldsymbol{\pi}_c\}_{c=1}^{C}$

$$
\begin{aligned}
&p(\{\boldsymbol{z}_{c(i)}\}_{i=1}^{N}, \boldsymbol{\upsilon}) = p(\boldsymbol{\upsilon}) \prod_i \int_{\boldsymbol{\pi}_{c(i)}} p(\boldsymbol{z}_{c(i)} | \boldsymbol{\pi}_{c(i)}) p(\boldsymbol{\pi}_{c(i)} | \boldsymbol{\upsilon}) \mathrm{d}\boldsymbol{\pi}_{c(i)} \\
&= \prod_{k=1}^{K} p\left(\upsilon_k | \{\upsilon_l\}_{l=1}^{k-1}\right) \times \prod_{c=1}^{C} \prod_{k=1}^{K} \frac{\Gamma(\gamma_c)}{\Gamma(\gamma_c + N_c)} \times \\
&\left(\sum_{q_{ck}=1}^{N_{ck}} \begin{bmatrix} N_{ck} \\ q_{ck} \end{bmatrix} (\gamma_c \upsilon_k)^{q_{ck}}\right) \times \left(\sum_{r_{ck}=1}^{\overline{N}_{ck}} \begin{bmatrix} \overline{N}_{ck} \\ r_{ck} \end{bmatrix} (\gamma_c - \gamma_c \upsilon_k)^{r_{ck}}\right),
\end{aligned}
$$

② Introducing auxiliary variables

$$
\begin{aligned}
&p\left(\upsilon_k | -\right) \propto \upsilon_k^{\lambda + \sum_c q_{ck} - 1} \bar{\upsilon}_k^{\sum_c l_{ck}} \mathbb{I}(\beta_{k+1} \le \beta_k \le \beta_{k-1}), \\
&p(q_{ck} | \boldsymbol{q}^{-(ck)}, \boldsymbol{\upsilon}, -) \propto \begin{bmatrix} N_{ck} \\ q_{ck} \end{bmatrix} (\gamma_c \upsilon_k)^{q_{ck}}, \\
&p(r_{ck} | \boldsymbol{r}^{-(ck)}, \boldsymbol{\upsilon}, -) \propto \begin{bmatrix} \overline{N}_{ck} \\ r_{ck} \end{bmatrix} (\gamma_c - \gamma_c \upsilon_k)^{r_{ck}},
\end{aligned}
$$

Learning Strategy via Sampling

Sampling for dictionary atoms

$$p(\mathbf{D}_{pc}|-) \propto \mathcal{N}\left(\mathbf{D}_{pc}; \alpha_c \Omega \mathrm{diag}(\boldsymbol{z}_c)(\textstyle\sum_{i:c(i)=c} x_{ip}\boldsymbol{s}_i), \Omega\right),$$
$$p(\boldsymbol{s}_i|c(i), -) \propto \mathcal{N}\left(\boldsymbol{s}_i; \alpha_n \boldsymbol{\Lambda}_{c(i)} \widetilde{\mathbf{D}}_c^{\mathsf{T}} \boldsymbol{x}_i, \boldsymbol{\Lambda}_{c(i)}\right),$$
$$p(\alpha_c|-) \propto \mathrm{Gamma}\Big(\alpha_c; g + \tfrac{P}{2}\textstyle\sum_{i=1}^{N} \mathbb{I}(c(i)=c),$$
$$h + \tfrac{1}{2}\textstyle\sum_{i:c(i)=c} \|\boldsymbol{x}_i - \mathbf{D}_{pc}\mathrm{diag}(\boldsymbol{z}_c)\boldsymbol{s}_i\|_2^2\Big),$$

Advantages

1. In every sampling steps, obtained posteriors still subject to original distribution types
2. No assumptions or approximations are introduce in the sampling procedures

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- **4. Experimental Results**
- 5. Conclusions

Experimental Results

- **Evaluations on the real dataset NUS-WIDE[1]:**


(a) Cat


(b) Elephant


(c) Hawk


(d) Whale


(e) Zebra

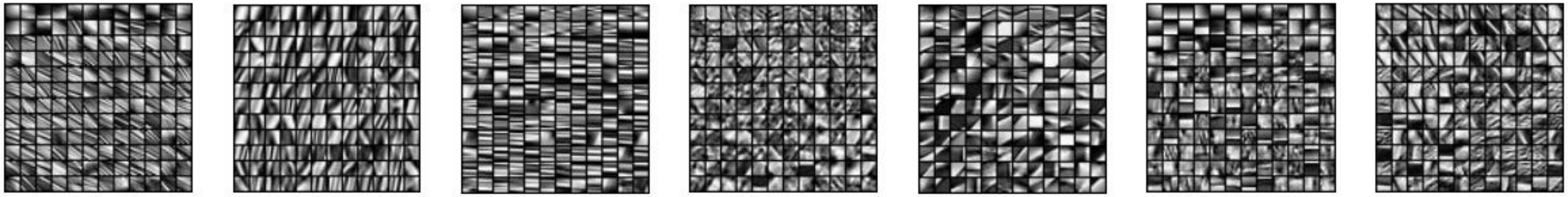
- **Experiment setup:**
  - **A subset involving 5 kinds of animals – *cat, elephant, hawk, whale, and zebra***
  - **The category number is unknown before inferencing the model**
  - **The parameters in the Bayesian model are initialized**

[1] T. S. Chua, et.al., "NUS-WIDE: A real-world web image database from National University of Singapore," ACM International Conference on Image and Video Retrieval, no. 48, pp. 1–9, Jul. 2009

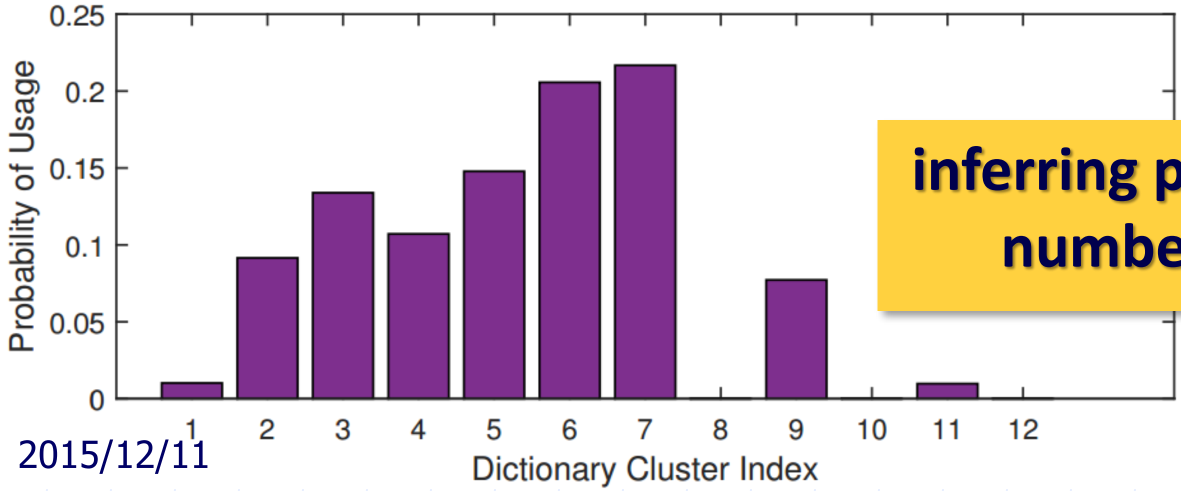
Experimental Results

Dictionary Learning with Training Images

The top seven frequently used dictionaries



The usage probability of the multiple learned dictionaries



inferring potential dictionary number automatically

Experimental Results

Image CS Recovery for Testing Dataset

One proxy for the multiple learned dictionaries

TABLE I

COMPARISONS OF THE RELATIVE RECONSTRUCTION ERROR WITH TESTING IMAGES FROM DIFFERENT CATEGORIES VERSUS M/P.

M/P \ Category		Cat Testing	Elephant Testing	Hawk Testing	Whale Testing	Zebra Testing
25%	BCS	0.2049	0.2012	0.1268	0.2066	0.2456
	BPFA	0.1168	0.1380	0.1015	0.1263	0.1531
	c-JDL	**0.1005**	**0.1190**	**0.0875**	**0.1025**	**0.1279**
35%	BCS	0.1651	0.1825	0.1138	0.1594	0.1950
	BPFA	0.1029	0.1184	0.0848	0.1120	0.1298
	c-JDL	**0.0857**	**0.1020**	**0.0752**	**0.0871**	**0.1118**
45%	BCS	0.1457	0.1643	0.1019	0.1463	0.1727
	BPFA	0.0996	0.1107	0.0798	0.1007	0.1209
	c-JDL	**0.0785**	**0.0933**	**0.0697**	**0.0801**	**0.1035**
55%	BCS	0.1254	0.1643	0.1019	0.1463	0.1588
	BPFA	0.0900	0.1072	0.0759	0.0988	0.1138
	c-JDL	**0.0666**	**0.0791**	**0.0590**	**0.0676**	**0.0948**

Outlines

- 1. Research Background
- 2. Nonparametric Bayesian Model for Multiple Dictionary Learning
- 3. Learning Strategy Based on Sampling
- 4. Experimental Results
- **5. Conclusions**

Conclusions

- **A nonparametric Bayesian approach to jointly learning multiple dictionaries for multiple image sources.**

- By introducing DP and HBP, the proposed model can infer the appropriate dictionary number from the image data and characterize the dictionary correlations

- An accurate inference engine for our model by utilizing an efficient Gibbs sampler in combination with a slice sampler based on auxiliary variables

- Proposed method shows performance enhancements over the existing Bayesian learning approach to reconstructing images from multiple sources

Thank you!