Bayesian Group Sparse Learning for Nonnegative Matrix Factorization

Jen-Tzung Chien and Hsin-Lung Hsieh

Department of Electrical and Computer Engineering
National Chiao Tung University, Hsinchu, Taiwan 30010, ROC
jtchien@ieee.org

Abstract

Nonnegative matrix factorization (NMF) is developed for parts-based representation of nonnegative data with the sparseness constraint. The degree of sparseness plays an important role for model regularization. This paper presents Bayesian group sparse learning for NMF and applies it for single-channel source separation. This method establishes the common bases and individual bases to characterize the shared information and residual noise in observed signals, respectively. Laplacian scale mixture distribution is introduced for sparse coding given a sparseness control parameter. A Markov chain Monte Carlo procedure is presented to infer two groups of parameters and their hyperparameters through a sampling procedure based on the conditional posterior distributions. Experiments on separating the single-channel audio signals into rhythmic and harmonic source signals show that the proposed method outperforms baseline NMF, Bayesian NMF and other group-based NMF in terms of signal-to-interference ratio.

Index Terms: Bayesian sparse learning, group sparsity, nonnegative matrix factorization, source separation

1. Introduction

Many problems in speech and audio processing can be tackled through matrix factorization. Different cost functions and constraints may lead to different factorized matrices. This procedure can identify the underlying sources from the mixed signals for blind source separation [2][9]. Nonnegative matrix factorization (NMF) is designed to find an approximate factorization $X \approx AS$ for a data matrix $X$ into a basis matrix $A$ and an encoding matrix $S$ which are all nonnegative. Different from principal component analysis and independent component analysis [5], NMF only allows additive combination due to the non-negative constraints on $A$ and $S$. NMF conducts a parts-based sparse representation where only a few components are active to encode input data. The sparseness constraint was imposed in objective function [8]. Such sparse coding is efficient and robust. However, controlling the sparseness or smoothness is crucial to achieve desirable sparse representation. Bayesian learning is beneficial to find the regularized NMF. Some related works have been proposed. In [1], Bayesian learning was performed for sparse representation of image data where Laplacian distribution was used as prior density for $\ell_1$-regularized optimization. In [11], group-based NMF was proposed to capture the intra-subject variations and the inter-subject variations in EEG signals. In [12], the group sparse NMF was proposed by minimizing the Itakura-Saito divergence between $X$ and $AS$. In [10], NMF was applied for drum source separation where the factorized components were partitioned into rhythmic and harmonic sources. No Bayesian learning was performed in [10][11][12]. Recently, a Bayesian NMF approach [3] was proposed for model selection and image reconstruction. This approach conducted model inference via the variational Bayes and Markov chain Monte Carlo (MCMC). In [13], a Bayesian NMF with Gamma priors for source signals and mixing coefficients was proposed through a MCMC algorithm. In [14], the Bayesian NMF with Gaussian likelihood and exponential priors was implemented for image feature extraction where the posterior distributions were approximated by Gibbs sampling procedure. All these methods [3][13][14] performed Bayesian NMF but without involving sparse and group learning.

In this paper, we present a new group-based NMF where the groups of common bases and individual bases [6] are estimated for blind separation of rhythmic sources and harmonic sources, respectively. Bayesian sparse learning is developed by introducing the Laplacian scale mixture distributions as the priors for two groups of encoding coefficients. Gamma priors are used to represent two groups of basis components. Accordingly, the Bayesian group sparse (BGS) learning is performed for BGS-NMF. A MCMC algorithm is derived to infer BGS-NMF parameters and hyperparameters. In the experiments, the proposed BGS-NMF is evaluated and compared with other NMF methods for single-channel separation of audio signals into rhythmic signals and harmonic signals.

2. Nonnegative Matrix Factorization

NMF is a linear model where the observed signals, factorized signals and source signals are assumed to be non-negative. Given a data matrix $X$, NMF estimates two
factorized matrices $A = \{a_{ij}\}$ and $S = \{s_{jk}\}$ by minimizing the reconstruction error between $X$ and $AS$. In [8], the sparseness constraint was imposed on minimizing the regularized Euclidean error function given by $\|X - AS\|^2 + \gamma_1 \sum_i f(a_{ij}) + \gamma_2 \sum_j f(s_{jk})$ where $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ are regularization parameters. Different sparseness measures could be used, e.g. $f(s_{jk}) = |s_{jk}|$, $f(s_{jk}) = s_{jk}$, $f(s_{jk}) = s_{jk} \log(s_{jk})$. In what follows, we survey several extensions of NMF.

2.1. Nonnegative matrix partial co-factorization

In [15], the nonnegative matrix partial co-factorization (NMPFC) was proposed for rhythmic source separation. Given the magnitude spectrogram as input data matrix $X$, NMPFC decomposes the music signal into the drum part and the residual part by $X = A_r S_r + A_h S_h$ with the factorized matrices for rhythmic sources $S_r$ and for harmonic sources $S_h$ based on common bases $A_r$ and for harmonic sources $S_h$ based on individual bases $A_h$, respectively. The prior knowledge from drum-only signal $Y \approx A_r S_r$, given the same rhythmic bases $A_r$ is incorporated in Euclidean error function $\|X - A_r S_r - A_h S_h\|^2 + \gamma \|Y - A_r S_r\|^2$ where $\gamma$ is a tradeoff between two errors. The joint minimization of reconstruction errors due to $X$ and $Y$ is performed.

In [10], the mixed signals were divided into $L$ segments. Each segment $X^{(l)}$ is decomposed into common and individual parts which reflect rhythmic and harmonic sources, respectively. The common bases $A_r$ are shared for different segments due to high temporal repeatability in rhythmic sources. The individual bases $A_h^{(l)}$ are separate for individual segment $l$ due to the changing frequency and low temporal repeatability. The objective function consists of a weighted Euclidean error function $\sum_l \omega^{(l)} \|X^{(l)} - A_r S_r^{(l)} - A_h^{(l)} S_h^{(l)}\|^2$ and the regularization terms of two bases $\|A_r\|^2$ and $\|A_h^{(l)}\|^2$ where $\{\omega^{(l)}, S_r^{(l)}, S_h^{(l)}\}$ denotes the segment-dependent weights, common encoding matrix and individual encoding matrix, respectively. This is a NMPFC for $L$ segments.

2.2. Group-based NMF

In [11], the group-based NMF (GNMF) was developed by conducting group analysis and constructing the task-related bases for intra-subject variations and inter-subject variations which reflect the response variations of a subject in different trials and those between different subjects, respectively. Given $L$ subjects or segments, the $l$-th segment is generated by $X^{(l)} = A_r^{(l)} S_r^{(l)} + A_h^{(l)} S_h^{(l)}$ where $A_r^{(l)}$ denotes the task-related common bases capturing the intra and inter-subject variations and $A_h^{(l)}$ denotes the task-independent bases reflecting the individual residual information. In general, different common bases $A_r^{(l)}$ are close together, but the individual bases $A_h^{(l)}$ are not. The objective function of GNMF is composed of five terms including $\sum_l \|X^{(l)} - A_r^{(l)} S_r^{(l)} - A_h^{(l)} S_h^{(l)}\|^2$, $\sum_l \|A_r^{(l)}\|^2$, $\sum_l \|A_h^{(l)}\|^2$, $\sum_l \|A_h^{(l)} - A_h^{(m)}\|^2$, and $-\sum_l \sum_m \|A_h^{(l)} - A_h^{(m)}\|^2$ where the second and the third terms are regularization functions, the fourth term enforces the distance between different common bases to be small, and the fifth term enforces the distance between different individual bases to be large.

3. Bayesian Group Sparse NMF

Previous NMF methods [10][11][15] were developed to extract task-specific nonnegative factors, but they did not consider the uncertainty of model parameters and did not control the sparsity of encoding coefficients. Here, we present a Bayesian group sparse learning for NMF (BG-NMF) and apply it for single-channel source separation.

3.1. Model construction

To deal with audio signal separation, we first calculate the magnitude spectrogram of audio signal and chunk the nonnegative data matrix $X$ into $\{X^{(l)}\}$. The $l$th segment is assumed to be generated by a noisy model

$$X^{(l)} = A_r S_r^{(l)} + A_h^{(l)} S_h^{(l)} + E^{(l)}$$

where the common bases $A_r$ are shared for different $X^{(l)}$, the individual bases $A_h^{(l)}$ are segment-dependent, and $E^{(l)}$ denotes the noise matrix of a segment. The sparseness constraint is imposed on two groups of encoding coefficients $\{S_r^{(l)}, S_h^{(l)}\}$. We assume that the encoding coefficients of rhythmic sources $S_r^{(l)}$ and harmonic sources $S_h^{(l)}$ are independent, but the dependencies between encoding coefficients within each group are allowed. Assuming noise signals are Gaussian distributed with zero means and diagonal variances $\{[\sigma^{(l)}_i]_i\}_{l=1}^L$, the likelihood function $p(X^{(l)} = \{X_{ik}^{(l)}\}|\Theta^{(l)})$ is expressed by

$$\prod_{l=1}^L \prod_{k=1}^K \mathcal{N}(X_{ik}^{(l)} \mid [A_r S_r^{(l)} + A_h^{(l)} S_h^{(l)}]_{ik}, \sigma^{(l)}_i)^2$$

BG-NMF model is accordingly constructed with parameters $\Theta^{(l)} = \{A_r, A_h^{(l)}, S_r^{(l)}, S_h^{(l)}, \{[\sigma^{(l)}_i]^2\}\}$.

3.2. Bayesian learning and sparse prior

From Bayesian perspective, the uncertainties of BG-NMF parameters, expressed by prior densities, are considered to assure model regularization. We seek the task-related common bases for different data segments. Also, we estimate the individual bases to reflect the unique characteristics in each segment. Sparsity control is enforced in encoding coefficients. Here, the priors of two groups of bases are specified by Gamma distributions

$$p(A_r) = \prod_{l=1}^L \prod_{j=1}^J \text{Gam}(a_{rij}|\alpha_{rj}, \beta_{rj})$$

$$p(A_h^{(l)}) = \prod_{l=1}^L \prod_{j=1}^J \text{Gam}(a_{hij}|\alpha_{hij}, \beta_{hij})$$

$$p(S_r^{(l)}) = \prod_{l=1}^L \prod_{j=1}^J \text{Gam}(s_{rij}|\alpha_{sij}, \beta_{sij})$$

$$p(S_h^{(l)}) = \prod_{l=1}^L \prod_{j=1}^J \text{Gam}(s_{hij}|\alpha_{sij}, \beta_{sij})$$
where $\Phi_a^{(t)} = \{ \{ \alpha_{ij}, \beta_{ij} \}, \{ \alpha_{hj}, \beta_{hj} \} \}$ denotes the hyperparameters and $\{ J_r, J_h \}$ denote the numbers of common bases and individual bases, respectively. We use the truncated Gamma priors where the distributions of negative bases are forced to zero. Importantly, we control the sparsity of coefficients by using the prior density based on the Laplacian scale mixture (LSM) distribution [7]. The LSM of encoding coefficient of common basis is constructed by $s_{rjk}^{(t)} = (\lambda^{(t)}_{rjk} - 1) u_{rjk}$ where $u_{rjk}$ has Laplacian distribution $p(u_{rjk}) = \frac{1}{2} \exp\{-|u_{rjk}|\}$ with scale 1 and $\lambda^{(t)}_{rjk}$ is a positive continuous mixture parameter. Variable $s_{rjk}$ has a Laplacian distribution $p(s_{rjk}|\lambda^{(t)}_{rjk}) = \frac{1}{2 \lambda^{(t)}_{rjk}} \exp\{-\lambda^{(t)}_{rjk} s_{rjk}\}$ with inverse scale $\lambda^{(t)}_{rjk}$ where $s_{rjk} \geq 0$ is considered. Assuming $p(\lambda^{(t)}_{rjk}) = \text{Gam}(\lambda^{(t)}_{rjk}; \gamma^{(t)}_{rjk}, \beta^{(t)}_{rjk})$, the marginal distribution of an encoding coefficient is calculated by [7]

$$p(A_h^{(t)}) = \prod_{t=1}^{T} \prod_{j=1}^{J_h} \text{Gam}(a_{hj}^{(t)}; \alpha_{hj}^{(t)}, \beta_{hj}^{(t)})$$

(4)

This LSM distribution is obtained by adopting the property that Gamma is the conjugate prior for Laplacian distribution. This distribution is sparse and even sparser than Laplacian distribution. Figure 1 compares LSM, Laplacian and Gaussian distributions. LSM is sharpest among these distributions. Again, we use a truncated LSM prior for nonnegative coefficient $s_{rjk}^{(t)}$. Similarly, the sparse prior of encoding coefficient $s_{hjk}$ can be expressed by hyperparameter $\lambda^{(t)}_{hjk}$ or $s_{hjk}$. We form the hyperparameters $\Phi_s^{(t)} = \{ \Phi_a^{(t)}, \Phi_s^{(t)} \} = \{ \lambda^{(t)}_{hjk}, \gamma_{hjk} \}$. 3.3. Model inference

In BGS-NMF framework, the posterior distribution of parameters and hyperparameters $p(\Theta, \Phi|X)$ is not analytically tractable. We develop a MCMC algorithm for approximate inference through iteratively generating samples according to the posterior distribution. This algorithm converges by those samples. Here, the segment index $l$ is neglected in the derivation. At each iteration $t$, we sequentially sample coefficient $s_{rjk}^{(t+1)}$ by using the conditional posterior distribution [13]

$$p(s_{rjk}|X_{(1:t)}, \Theta_{rjk}^{(t)}, \Phi_{rjk}^{(t)}) \propto$$

$$p(X_{(1:t)}|s_{rjk}, \Theta_{rjk}^{(t)}, \Phi_{rjk}^{(t)}) p(s_{rjk})$$

(6)

where $X_{(1:t)} = \\{ [X]_{ik}\}^{t}_{i=1}$, $\Theta_{rjk} = \{ \theta_{rjk} \}$ and $\Phi_{rjk} = \{ \Phi_{rjk} \}$ in each sampling, we use the preceding coefficients $s_{rjk}^{(t)}$ at new iteration $t+1$ and subsequent coefficients $s_{rjk}^{(t+1)}$, $s_{rjk}^{(t+1)} = \{ [s_{rjk}]_{m=0:j+1} \}$ at current iteration $t$. The likelihood function can be derived as a Gaussian distribution

$$p(X_{(1:t)}|s_{rjk}, \Theta_{rjk}^{(t)}) \propto \exp\left\{ -\frac{(s_{rjk} - \mu_{rjk}^{(t)})^2}{2\sigma_{rjk}^{(t)}^2} \right\}$$

(7)

where the Gaussian parameters $\mu_{rjk}^{(t)}, \sigma_{rjk}^{2}$ are derived similar to [13]. Given the Gaussian likelihood and LSM prior, the conditional posterior distribution is calculated by (6). However, this conditional posterior is not a usual distribution. Its sampling requires the Metropolis-Hastings (M-H) algorithm [4] where the best instrumental distribution $q(s_{rjk})$ is selected to fit the target distribution at each sampling iteration $t$. The instrumental distribution is derived as a truncated and nonnegative-valued Gaussian distribution $\mathcal{N}_+^{(t)}(s_{rjk}|\mu_{s_{rjk}}^{(t)}, \sigma_{s_{rjk}}^{(t)})$ where the mode of instrumental distribution $\mu_{s_{rjk}}^{(t)}$ is derived by finding the root of a second-order equation of $s_{rjk}$ appearing in the exponent term of conditional posterior distribution [13]. Similarly, the sampling of encoding coefficient $s_{hjk}^{(t)}$ is performed according to a truncated Gaussian distribution $g(s_{hjk}) = \mathcal{N}_+(s_{hjk}|\mu_{s_{hjk}}^{(t)}, \sigma_{s_{hjk}}^{(t)})$ derived by M-H algorithm.

To perform the Gibbs sampling for common basis $a_{rij}^{(t+1)}$ or individual basis $a_{hij}^{(t+1)}$, we calculate the conditional posterior distribution which combines a likelihood function and a Gamma prior in (3) or (4). The M-H algorithm is applied to find the instrumental distributions $g(a_{rij})$ and $g(a_{hij})$. Further, the inverse of noise variance $\sigma^{-2}_{(t+1)}$ is sampled according to a conditional posterior which is derived as a Gamma distribution due to a Gaussian likelihood and a Gamma prior [13]. In addition to model parameters $\Theta^{(t)}$, the Gamma hyperparameters $\Phi_a^{(t)}$ and LSM hyperparameters $\Phi_s^{(t)}$ are also sampled for
BGS-NMF by following the M-H algorithm. Notably, different from [13], the proposed BGS-NMF conducts the group sparse learning based on the LSM distribution. Common bases $A_r$ are shared for different data segments $l$. The group sparsity performs well in our experiments.

4. Experiments

We evaluated BGS-NMF for single-channel source separation with two audio source signals. The rhythmic and harmonic source signals were sampled from http://www.free-scores.com/index_uk.php3 and http://www.freesound.org/. Six sets of source signals were collected as follows: (1) bass+piano, (2) drum+guitar, (3) drum+violin, (4) cymbal+organ, (5) drum+saxophone, and (6) cymbal+singing, which contained different combinations of rhythmic and harmonic sources. We used different mixing matrices for six datasets. Each audio signal was 21 seconds long with a 44,100-Hz sampling rate and 16-bit resolution. The magnitude of FFT of audio signal was extracted every 1024 samples with 512 samples in frame overlapping. Each mixed signal was equally chopped into $L$ segments with length of 3 seconds for source separation. Sufficient rhythmic signal existed within a segment. The number of common bases and individual bases was set to be 15 and 10, respectively. System performance was insensitive to these numbers. We performed 1000 Gibbs sampling iterations. The separation performance was evaluated by the signal-to-interference ratio (SIR) which was averaged over six datasets. The noise was defined as the difference between original and reconstructed source signals. We carried out baseline NMF [8], Bayesian NMF (BNMF) [13], GNMF [11] and proposed BGS-NMF under consistent experimental setup. Table 1 shows the experimental results. The SIRs of reconstructed rhythmic and harmonic signals using BGS-NMF are measured as 8.3997 dB and 8.1295 dB which are better than 4.8658 dB and 4.6085 dB using BNMF and 5.6343 dB and 5.7065 dB using GNMF, respectively. BGS-NMF attains the best separation performance among four different NMFs. The waveforms and spectrograms of the original signals, mixed signals and demixed signals based on BGS-NMF for six datasets are accessible at http://chien.csie.ncku.edu.tw/bgs-nmf.

<table>
<thead>
<tr>
<th></th>
<th>NMF</th>
<th>BNMF</th>
<th>GNMF</th>
<th>BGS-NMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>3.70</td>
<td>3.32</td>
<td>4.87</td>
<td>4.61</td>
</tr>
<tr>
<td>H</td>
<td>5.63</td>
<td>5.71</td>
<td>7.13</td>
<td>8.40</td>
</tr>
</tbody>
</table>

Table 1: Comparison of SIRs (in dB) of the reconstructed rhythmic (R) and harmonic (H) sources using different NMFs.

5. Conclusions

This paper has presented the Bayesian group sparse learning and applied it for single-channel source separation. The basis vectors in NMF were grouped into two partitions. One was the common bases for exploring the inter-segment characteristics and the other was the individual bases for representing the intra-segment information. LSM distribution was introduced to express the sparse encoding coefficients for two groups of basis vectors. The MCMC algorithm was developed for approximate inference of model parameters and hyperparameters. In the experiments, we implemented the proposed BGS-NMF for underdetermined source separation and found that BGS-NMF outperformed other NMFs under different kinds of rhythmic and harmonic sources and mixing conditions.

6. References