BAYESIAN SINGING-VOICE SEPARATION

Guan-Xiang Wang, Po-Kai Yang, Chung-Chien Hsu and Jen-Tzung Chien

Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan
gxwang@chien.cm.nctu.edu.tw, {niceallen.cm01g, chien.cm97g, jtchien}@nctu.edu.tw

ABSTRACT

This paper presents a Bayesian nonnegative matrix factorization (NMF) approach to extract singing voice from background music accompaniment. Using this approach, the reconstruction error based on NMF is represented by a Poisson distribution and the NMF parameters, consisting of basis and weight matrices, are characterized by the exponential priors. A variational Bayesian expectation-maximization algorithm is developed to learn variational parameters and model parameters for monaural source separation. A clustering algorithm is performed to establish two groups of bases: one is for singing voice and the other is for background music. Model complexity is controlled by adaptively selecting the number of bases for different mixed signals according to the variational lower bound.

1. INTRODUCTION

This paper presents a new model-based singing-voice separation. The novelties of this paper are twofold. The first one is to develop Bayesian approach to unsupervised singing-voice separation. Model uncertainty is compensated to improve the performance of source separation of vocal signal and background accompaniment signal. Number of bases is adaptively determined from the mixed signal according to the variational lower bound of the marginal likelihood over NMF basis and weight matrices. The second one is the theoretical contribution in Bayesian NMF. We construct a new Bayesian NMF where the modeling error in NMF is drawn from Poisson distribution and the model parameters are characterized by exponential distributions.

2. NEW BAYESIAN NONNEGATIVE MATRIX FACTORIZATION

This study aims to find an analytical solution to full Bayesian NMF by considering all dependencies of variational lower bound on regularization parameters. Regularization parameters are optimally estimated.

2.1 Bayesian Objectives

We adopt the Poisson distribution as likelihood function and the exponential distribution as *conjugate prior* for NMF parameters B_{mk} and W_{kn} with hyperparameters λ^b_{mk} and λ^w_{kn}, respectively. Maximum *a posteriori* (MAP) estimates of parameters $\Theta = \{\mathbf{B}, \mathbf{W}\}$ are obtained by maximizing the posterior distribution or minimizing $-\log p(\mathbf{B}, \mathbf{W}|\mathbf{X})$ which is arranged as a regularized KL divergence between $\mathbf{X}$ and $\mathbf{BW}$

$$D_{\mathrm{KL}}(\mathbf{X}||\mathbf{BW}) + \sum_{m,k} \lambda^b_{mk} B_{mk} + \sum_{k,n} \lambda^w_{kn} W_{kn} \tag{1}$$

where the terms independent of B_{mk} and W_{kn} are treated as constants. Notably, the regularization terms (2nd and 3rd terms) in this objective are nonnegative and seen as the ℓ_1 regularizers [2] which are controlled by hyperparameters $\{\lambda^b_{mk}, \lambda^w_{kn}\}$. These regularizers impose sparseness in the estimated MAP parameters.

However, MAP estimates are seen as point estimates. The randomness of parameters is not considered in model construction. To conduct full Bayesian treatment, BNMF is developed by maximizing the marginal likelihood $p(\mathbf{X}|\Theta)$ over latent variables $\mathbf{Z}$ as well as NMF parameters $\{\mathbf{B}, \mathbf{W}\}$

$$\int \sum_{\mathbf{Z}} p(\mathbf{X}|\mathbf{Z}, \mathbf{B}, \mathbf{W}) p(\mathbf{Z}|\mathbf{B}, \mathbf{W}) p(\mathbf{B}, \mathbf{W}|\Theta) d\mathbf{B} d\mathbf{W} \tag{2}$$

and estimating the sparsity-controlled hyperparameters or regularization parameters $\Theta = \{\lambda^b_{mk}, \lambda^w_{mk}\}$. The resulting evidence function is meaningful to act as an objective for model selection which balances the tradeoff between data fitness and model complexity [1]. In the singing-voice separation based on NMF, this objective is used to judge which number of bases K should be selected.

2.2 Variational Bayesian Inference

The variational Bayesian expectation-maximization (VB-EM) algorithm is developed to implement Poisson-Exponential BNMF. VB-EM algorithm applies the Jensen's inequality and maximizes the lower bound of the logarithm of marginal likelihood

$$\log p(\mathbf{X}|\Theta) \geq \int \sum_{\mathbf{Z}} q(\mathbf{Z}, \mathbf{B}, \mathbf{W}) \log \frac{p(\mathbf{X}, \mathbf{Z}, \mathbf{B}, \mathbf{W}|\Theta)}{q(\mathbf{Z}, \mathbf{B}, \mathbf{W})} \times d\mathbf{B} d\mathbf{W} = \mathbb{E}_q[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{B}, \mathbf{W}|\Theta)] + H[q(\mathbf{Z}, \mathbf{B}, \mathbf{W})] \tag{3}$$

where $H[\cdot]$ is an entropy function. The factorized variational distribution $q(\mathbf{Z}, \mathbf{B}, \mathbf{W}) = q(\mathbf{Z})q(\mathbf{B})q(\mathbf{W})$ is

assumed to approximate the true posterior distribution $p(\mathbf{Z}, \mathbf{B}, \mathbf{W}|\mathbf{X}, \mathbf{\Theta})$.

2.2.1 VB-E Step

In VB-E step, a general solution to variational distribution q_j of an individual latent variable $j \in \{\mathbf{Z}, \mathbf{B}, \mathbf{W}\}$ is obtained by [1]

$$\log \hat{q}_j \propto \mathbb{E}_{q_{(i \neq j)}}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{B}, \mathbf{W}|\Theta)]. \tag{4}$$

Given the variational distributions defined by

$$\begin{aligned} q(B_{mk}) &\propto \text{Gam}(B_{mk}; \alpha^b_{mk}, \beta^b_{mk}) \\ q(W_{kn}) &\propto \text{Gam}(W_{kn}; \alpha^w_{kn}, \beta^w_{kn}) \\ q(Z_{mkn}) &\propto \text{Mult}(Z_{mkn}; P_{mkn}) \end{aligned} \tag{5}$$

the variational parameters $\{\alpha^b_{mk}, \beta^b_{mk}, \alpha^w_{kn}, \beta^w_{kn}, P_{mkn}\}$ in three distributions are estimated by

$$\begin{aligned} \hat{\alpha}^b_{mk} &= 1 + \sum_n \langle Z_{mkn} \rangle, \;\; \hat{\beta}^b_{mk} = \left(\sum_n \langle W_{kn} \rangle + \lambda^b_{mk} \right)^{-1} \\ \hat{\alpha}^w_{kn} &= 1 + \sum_m \langle Z_{mkn} \rangle, \;\; \hat{\beta}^w_{kn} = \left(\sum_k \langle B_{mk} \rangle + \lambda^w_{kn} \right)^{-1} \\ \hat{P}_{mkn} &= \frac{\exp(\langle \log B_{mk} \rangle + \langle \log W_{kn} \rangle)}{\sum_j \exp(\langle \log B_{mj} \rangle + \langle \log W_{jn} \rangle)} \end{aligned} \tag{6}$$

where the expectation function $\mathbb{E}_q[\cdot]$ is replaced by $\langle \cdot \rangle$ for simplicity. By substituting the variational distribution into Eq. (3), the variational lower bound is obtained by

$$\begin{aligned} \mathcal{B}_L = & - \sum_{m,n,k} \langle B_{mk} \rangle \langle W_{kn} \rangle \\ & + \sum_{m,n} (-\log \Gamma(X_{mn} + 1) - \sum_k \langle Z_{mkn} \rangle \log \hat{P}_{mkn}) \\ & + \sum_{m,k} \langle \log B_{mk} \rangle \sum_n \langle Z_{mkn} \rangle + \sum_{k,n} \langle \log W_{kn} \rangle \sum_m \langle Z_{mkn} \rangle \\ & + \sum_{m,k} (\log \lambda^b_{mk} - \lambda^b_{mk} \langle B_{mk} \rangle) + \sum_{k,n} (\log \lambda^w_{kn} - \lambda^w_{kn} \langle W_{kn} \rangle) \\ & + \sum_{m,k} (-(\hat{\alpha}^b_{mk} - 1)\Psi(\hat{\alpha}^b_{mk}) + \log \hat{\beta}^b_{mk} + \hat{\alpha}^b_{mk} + \log \Gamma(\hat{\alpha}^b_{mk})) \\ & + \sum_{k,n} (-(\hat{\alpha}^w_{kn} - 1)\Psi(\hat{\alpha}^w_{kn}) + \log \hat{\beta}^w_{kn} + \hat{\alpha}^w_{kn} + \log \Gamma(\hat{\alpha}^w_{kn})) \end{aligned} \tag{7}$$

where $\Psi(\cdot)$ is the derivative of the log gamma function, and is known as a digamma function.

2.2.2 VB-M Step

In VB-M step, the optimal regularization parameters $\Theta = \{\lambda^b_{mk}, \lambda^w_{kn}\}$ are derived by maximizing Eq. (7) with respect to Θ and yielding

$$\begin{aligned} \frac{\partial \mathcal{B}_L}{\partial \lambda^b_{mk}} &= \frac{1}{\lambda^b_{mk}} - \langle B_{mk} \rangle + \frac{\partial \log \beta^b_{mk}}{\partial \lambda^b_{mk}} = 0 \\ \frac{\partial \mathcal{B}_L}{\partial \lambda^w_{kn}} &= \frac{1}{\lambda^w_{kn}} - \langle W_{kn} \rangle + \frac{\partial \log \beta^w_{kn}}{\partial \lambda^w_{kn}} = 0. \end{aligned} \tag{8}$$

Accordingly, the solution to BNMF hyperparameters is derived by solving a quadratic equation where nonnegative constraint is considered to find positive values of hyperparameters by

$$\begin{aligned} \hat{\lambda}^b_{mk} &= \frac{1}{2} \left(- \sum_n \langle W_{kn} \rangle + \sqrt{(\sum_n \langle W_{kn} \rangle)^2 + 4 \frac{\sum_n \langle W_{kn} \rangle}{\langle B_{mk} \rangle}} \right) \\ \hat{\lambda}^w_{kn} &= \frac{1}{2} \left(- \sum_m \langle B_{mk} \rangle + \sqrt{(\sum_m \langle B_{mk} \rangle)^2 + 4 \frac{\sum_m \langle B_{mk} \rangle}{\langle W_{kn} \rangle}} \right) \end{aligned} \tag{9}$$

where $\langle B_{mk} \rangle = \alpha^b_{mk} \beta^b_{mk}$ and $\langle W_{kn} \rangle = \alpha^w_{kn} \beta^w_{kn}$ are obtained as the means of gamma distributions.

2.3 Poisson-Exponential Bayesian NMF

In this study, total number of basis vectors K is adaptively selected for individual mixed signal according to the variational lower bound in Eq. (7) with the converged variational parameters $\{\hat{\alpha}^b_{mk}, \hat{\beta}^b_{mk}, \hat{\alpha}^w_{kn}, \hat{\beta}^w_{kn}, \hat{P}_{mkn}\}$ and model parameters $\{\hat{\lambda}^b_{mk}, \hat{\lambda}^w_{kn}\}$.

Considering the pairs of likelihood function and prior distribution in NMF, the proposed method is also called the Poisson-Exponential BNMF.

2.4 Unsupervised Singing-Voice Separation

We implemented the unsupervised singing-voice separation where total number of bases (K) and the grouping of these bases into vocal source and music source were both learned from test data in an unsupervised way. We conduct NMF-based clustering for the proposed BNMF method. To do so, we transformed the basis vectors $\mathbf{B}$ into Mel-scaled spectrum to form the Mel-scaled basis matrix. ML-NMF was applied to factorize this Mel-scaled basis matrix into two matrices $\tilde{\mathbf{B}}$ of size N-by-2 and $\tilde{\mathbf{W}}$ of size 2-by-K. The soft mask scheme based on Wiener gain was applied to smooth the separation of $\mathbf{B}$ into basis vectors for vocal signal and music signal. This same soft mask was performed for the separation of mixed signal X into vocal signal and music signal based on the K-means clustering and NMF clustering. Finally, the separated singing voice and music accompaniment signals were obtained by the overlap-and-add method using the original phase.

3. CONCLUSIONS

We proposed a new unsupervised Bayesian nonnegative matrix factorization approach to extract the singing voice from background music accompaniment and illustrated the novelty on an analytical and true optimum solution to the Poisson-Exponential BNMF. Through the VB-EM inference procedure, the proposed method automatically selected different number of bases to fit various experimental conditions.

4. REFERENCES

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science, 2006.

[2] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.