ONLINE BAYESIAN LEARNING FOR DYNAMIC SOURCE SEPARATION
Hsin-Lung Hsieh and Jen-Tzung Chien
Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan, Taiwan 70101, ROC
{hlhsieh,chien}@chien.csie.ncku.edu.tw

ABSTRACT

Independent component analysis (ICA) is a popular approach for blind source separation when the source signals are stationary with fixed distribution functions. However, the source signals are nonstationary in real-world applications, e.g., the source signals may abruptly appear or disappear or even the number of sources may be changed by time. This study presents a nonstationary ICA for dynamic source separation through an online Bayesian learning procedure. In this procedure, we capture the evolved statistics of independent sources from the online observed signals. The mixing matrix is incrementally compensated at each frame and continuously propagated to the next frame. A variational Bayesian algorithm is established to estimate the nonstationary ICA parameters. The number of latent sources is automatically determined at each frame. In the experiments, the proposed method effectively recovers the source speech signals from different speakers in presence of different mixing scenarios.

Index Terms— Signal processing, separation, variational methods, Bayes procedures

1. INTRODUCTION

Independent component analysis (ICA) is developed for recovering the original independent source signals under the situations that we only observe the mixed signals and the actual mixing process is unknown. Let \( x = [x_1, \ldots, x_N]^T \) denote the observations which are linear mixtures of \( M \) statistically independent sources \( s = [s_1, \ldots, s_M]^T \) by \( x = As \). In traditional ICA methods, a demixing matrix \( W \) is found by minimizing the mutual information between the demixed signals \( s = Wx \) which is designed for static mixing problem in presence of the stationary distributions. However, in real-world BSS applications, the source signals may involve various nonstationary scenarios such as a sudden presence or absence of source signals or a time-varying mixing of source signals. Correspondingly, the mixing matrix and the distributions of sources are changed by time. To solve such a complicated circumstance, we need to identify the number of latent sources and trace their probability density functions at each time moment.

There have been several methods proposed to cope with the situations of dynamic sources or compensate the variations of mixing system in nonstationary environments. In general, there are two scenarios in dynamic source separation. First, the source signals or the sensors may be moving. For this case, we need to compensate the variations of mixing matrix. In [6], a state transition matrix was adopted to capture the variations in mixing matrix when the speaker or microphone was moving but the number of sources was fixed. In the second scenario, the source signals may suddenly appear, disappear or be permuted at different time moments. For this case, we need to continuously estimate the time-varying source distributions or the number of source signals. The number of source signals was selected by an unsupervised learning method [3] using the automatic relevance determination (ARD) [10]. In [4], a hidden Markov model (HMM) was presented to capture the temporal information and merge the prior information of mixture coefficients for Bayesian ICA. The source distributions were dependent on the state of HMM. In [7], an indicator variable was used to indicate if the sources are active or inactive in a switching ICA. A special type of HMM was adopted to represent this scenario. The computational cost was high due to the incorporation of Markov chain in ICA procedure. In [8], an online VB learning was performed in ICA procedure where the source signal was seen as a time-dependent parameter and the mixing matrix was referred as a time-independent parameter. The source distributions were updated incrementally while the number of sources was assumed to be fixed. Such an approach could not tackle the case of abruptly appearing or disappearing source signal.

Regarding the ICA model inference, Bishop and Lawrence [1] presented the variational Bayesian (VB) learning instead of maximum likelihood learning. In [4], the source signal was modeled by the mixture of Gaussian model and the variational learning was applied for Bayesian ICA. In [11], an expectation consistent framework was exploited to perform the approximate inference and the resulting expectation-maximization (EM) optimization was shown to be efficient. These studies were not designed for dynamic source separation. To deal with the dynamic source separation from the perspective of Bayesian philosophy, we are motivated to develop the online Bayesian learning [2] to carry out the nonstationary ICA. An online Bayesian learning algorithm is presented to build a nonstationary ICA procedure for dynamic source separation where the mixing matrix and the distribution of source signals are jointly estimated at each time moment via an incremental VB estimation. The computational cost is reduced. A compensation parameter is introduced to adjust the mixing matrix. This parameter is updated continuously from online observed signals. The proposed method is validated by the experiments on different scenarios in presence of dynamic sources due to multiple speakers.

2. SURVEY OF RELATED WORKS

Standard ICA assumes that the observation data are linearly combined by statistically independent sources, which are identically distributed. However, in real-world applications, we encounter the circumstances of dynamic changes in source signals, e.g., the source signals may appear or disappear and the source variables may be moved. The temporal information of source signals should be carefully treated in ICA procedure. Choudrey and Roberts [5] incorporated an HMM in ICA method for representation of temporal changes in source signals. The data generation using the noisy ICA model is represented by...
\[ x_t = A s_t + \varepsilon_t \]

where the source signal \( s_t \), observation signal \( x_t \) and noise signal \( \varepsilon_t \) are time dependent. The \( m \)th source signal \( s_{mt} \) is modeled by a mixture of \( I_m \) Gaussians

\[
p(s_t | \theta) = \prod_{m=1}^{I_m} \sum_{n=1}^{N_m} \pi_{s_{mt}} N(s_{mt} | \mu_{s_{mt}}, \gamma_{s_{mt}}^{-1})
\]

where \( \theta = \{ \pi_{s_{mt}}, \mu_{s_{mt}}, \gamma_{s_{mt}} \} \) is the Gaussian mixture parameters and \( \pi_{s_{mt}} = p(q_{mt} = q_m) \) is a mixture parameter of choosing a Gaussian at state \( q_m \) with mean \( \mu_{s_{mt}} \) and precision \( \gamma_{s_{mt}} \). The temporal information of the sources is expressed by the observation probability at source states \( \{ q_{mt} \} \). Such an approach is helpful for detecting the dynamic change of source signals and capturing the nonstationary source variables for VB learning of ICA model. In addition, the standard ICA assumes that a set of fixed sources exist at the whole time series. The temporary absence or presence of source signal is not allowed. Hirayama et al. [7] proposed a switching ICA to overcome the circumstance that the status of the source signal \( s_{mt} \) may be active (presence) or inactive (absence) with time. The status was indicated by a switching variable \( z_{mt} \). The value \( z_{mt} = 0 \) implies an inactive source while the value \( z_{mt} = 1 \) means the source is active. The source signal in case of \( z_{mt} = 1 \) was modeled by a scaled-mixture of Gaussians [7].

A Markov process was used to characterize the temporal variations at each source. Nevertheless, these methods [5][7] conducted the batch learning strategy. The Markov states in \( M \) source signals were searched. The model parameters were estimated by the VB inference. The switching ICA was computationally demanding since the computation was closely related to the number of states in switching variables which was expanded exponentially by the number of sources \( M \). A single mixture matrix was estimated.

3. Dynamic Source Separation

3.1 Online Bayesian Learning

In a realistic signal separation system, the environmental conditions are changed naturally as the time moves on. A flexible ICA procedure should be established by relaxing the assumptions in standard ICA. This paper presents a new nonstationary Bayesian ICA (NB-ICA) algorithm by incrementally identifying the dynamic sources via estimating the ICA parameters and an online compensation parameter. The sufficient statistics from the previous frames is propagated and merged with the likelihood of current frame for adaptive signal separation. The mixing matrix and distributions of source signals are incrementally estimated instead of performing Markov chain for each source signal in switching ICA [5][7]. The computational load can be reduced accordingly. Our idea is to incrementally detect the status of source signals and estimate the corresponding distributions from online observation data \( \chi' = \{ X_1, X_2, \ldots, X_t \} \). A transformation function \( G_t(\cdot) \) is introduced to characterize the dynamics of mixing matrix and also compensate the variations in signal separation. At time frame \( t \), we estimate ICA parameter \( \theta^{(t)} \) and transformation parameter \( \eta^{(t)} \) by maximizing the posterior distribution using \( \chi' \). Since the number and the distribution of latent sources are varied with time, it is preferable to perform incremental signal separation by Bayesian framework. The problem turns out to maximize the product of a likelihood function of current frame with the mixed samples \( X_t = \{ \chi_t \} \) and a posterior distribution given the previous frames of mixed samples \( \chi_t^{-1} \) [2]. The posterior distribution \( p(\theta^{(t)}, \eta^{(t)} | \chi' ) \) is calculated through the stages of prediction and correction. Given the historical data \( \chi_t^{-1} \), the prediction of the parameters \( \{ \theta^{(t)}, \eta^{(t)} \} \) is obtained according to the probability

\[
p(\theta^{(t)}, \eta^{(t)} | \chi_t^{-1}) \text{ calculated by}
\]

\[
\int [p(\theta^{(t)}, \eta^{(t)} | \chi_t^{-1})] p(\theta^{(t)}, \eta^{(t)} | \chi_t^{-1}) d\theta^{(t)} d\eta^{(t)}.
\]

When a new frame \( X_t \) is observed, the prediction is corrected by

\[
p(\theta^{(t)}, \eta^{(t)} | X_t) = \frac{p(X_t | \theta^{(t)}, \eta^{(t)}) p(\theta^{(t)}, \eta^{(t)} | \chi_t^{-1})}{\int [p(X_t | \theta^{(t)}, \eta^{(t)}) p(\theta^{(t)}, \eta^{(t)} | \chi_t^{-1})] d\theta^{(t)} d\eta^{(t)}}.
\]

The maximum a posteriori (MAP) estimation in this NB-ICA is converted into the following two stages

\[
\theta^{(t)} = \arg \max_{\theta} p(X_t | \theta, \eta^{(t-1)}) p(\theta)
\]

\[
\eta^{(t)} = \arg \max_{\eta} p(X_t | \theta^{(t)}, \eta) p(\eta) \phi^{-1}. \tag{6}
\]

The parameters are sequentially estimated frame by frame by \((\theta^{(t)}, \eta^{(t)}) \rightarrow \cdots \rightarrow (\theta^{(T)}, \eta^{(T)})\). In this NB-ICA procedure, ICA parameters \( \theta^{(t)} \) are first estimated by (5) and then plugged into (6) to calculate the transformation parameter \( \eta^{(t)} \) for compensating the variations in ICA mixing matrix. The prior densities \( p(\theta) \) and \( p(\eta | \phi^{-1}) \) with hyperparameter \( \phi^{-1} \) are involved. Here, we use a conjugate prior to specify \( p(\eta | \phi^{-1}) \). A recursive Bayesian algorithm can be formulated for updating the posterior distribution, which belongs to the same distribution family as prior density and is expressed by \( p(\eta | \phi') \) with the updated hyperparameters \( \phi^{-1} \rightarrow \phi' \). Different from the batch training in ICA procedure of [5][7], we present an efficient approach to finding the status of each source signal and estimating the distributions of reconstructed source signals frame by frame. The proposed NB-ICA works for different scenarios of dynamic source separation. In what follows, we address the model construction and inference by using NB-ICA.

3.2 Model Construction

In this study, the signal frame is generated by a noisy ICA model as given in (1). Let \( X_t = \{ \chi_t \} \) denote \( N \times 1 \) mixed signal vectors. The signals are assumed to be mixed by a linear combination of \( M \) unknown source signals \( S^{(t)} = \{ s^{(t)} \} \) plus a noise \( \varepsilon_t \). The graphical representation of an online learning procedure is illustrated by Figure 1. Here, the prior density of the mixing matrix \( A^{(t)} \) is expressed by \( \prod_{m=1}^{M} \prod_{n=1}^{N} N(a_{mn}^{(t)} | 0, \sigma_m^{-1}) \) where each entry \( a_{mn}^{(t)} \) is an isotropic Gaussian with zero mean and precision \( \sigma_m^{-1} \). The precision \( \sigma_m^{-1} \) has a Gamma prior, i.e.

\[
p(\sigma_m^{-1}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m^{(0)}, \beta_m^{(0)}).
\]

The noise vector \( \varepsilon_t \) is assumed to be an isotropic Gaussian \( N(\varepsilon_t | 0, \beta_t^{-1} I_N) \) with zero mean and precision \( \beta_t^{-1} \). \( I_N \) is an \( N \times N \) identity matrix. We choose a Gamma prior for \( \beta_t^{-1} \), i.e. \( \text{Gamma}(\beta_t^{(0)} | a_{t, \beta}^{(0)}, b_{t, \beta}^{(0)}) \). The likelihood function of an observation signal \( x_t \) is yielded by
A mixture of $K$ Gaussians is adopted to represent the source signals $s^{(i)} = [s^{(i)}_1, \ldots, s^{(i)}_M]^T$.

$$p(s^{(i)} | \pi^{(i)}, \mu^{(i)}, \gamma^{(i)}) = \prod_{n=1}^M \sum_{k=1}^K \frac{1}{\gamma^{(n)}} \frac{1}{\sqrt{2\pi \gamma^{(n)}}} \exp\left(-\frac{1}{2\gamma^{(n)}} (s^{(i)}_n - \mu^{(i)}_k)^2\right)$$

(8)

where $\theta^{(i)} = \{\pi^{(i)}_k, \mu^{(i)}_k, \gamma^{(i)}_k\}$ contains the mixture weight, mean and precision. The parameter set is formed by $\theta^{(i)} = \{A^{(i)}, s^{(i)}, a^{(i)}, \theta^{(i)}\}$. The marginal likelihood $p(X_i)$ using NB-ICA model is obtained by

$$p(X_i) = \prod_{i=1}^T \int p(X_i | A^{(i)}, s^{(i)}, a^{(i)}, \theta^{(i)}) p(a^{(i)} \mid a_0^{(i)}, w^{(i)}) p(a_0^{(i)} \mid u_0^{(i)}, w^{(i)}) p(u^{(i)} \mid m_0^{(i)}, d^{(i)} A^{(i)} \mu^{(i)} d^{(i)} a^{(i)} d\theta^{(i)})$$

(9)

Interestingly, each entry of precision parameter $a^{(i)} = [a^{(i)}_1, \ldots, a^{(i)}_M]$ implies the relevance or the activity of a source signal in the mixing system. Finding the parameter $a^{(i)}$ is equivalent to conducting the automatic relevance determination (ARD) [10] for pruning the redundant source components automatically. If the estimated $a^{(i)}_m$ is very large, $a^{(i)} = [a^{(i)}_1, \ldots, a^{(i)}_M] \rightarrow 0$ is obtained and it implies that the $m$th source has high sparsity and could be pruned. This ARD parameter $a^{(i)}$ plays a similar role of the indicator variable $z^{(i)}$ in switching ICA [7].

This NB-ICA is computationally efficient since no Markov chain is involved in the ICA procedure. Importantly, the dynamics of mixing matrix are compensated through the incremental adaptation of $a^{(i)}$. To do so, we address the incremental adaptation of precision parameter $a^{(i)}$ of mixing matrix by the compensation parameter $\eta^{(i)}$. Here, we choose the conjugate prior - Wishart distribution with hyperparameters $\eta^{(-1)} = \{\nu^{(-1)}, A^{(-1)}\}$ for the scaling matrix of the precision parameter of Gaussian distribution

$$p(\eta^{(i)} | \phi^{(-1)} = \{\nu^{(-1)}, A^{(-1)}\}) = \frac{c(M, \nu^{(-1)})}{2} \frac{1}{\nu^{(-1)}} \frac{1}{A^{(-1)}} \left(\frac{\nu^{(-1)}-2}{2}\right)^{(M-2)/2} \exp\left[-\frac{1}{2\nu} \text{tr}(\nu^{-1} A^{-1} \eta^{(i)})\right]$$

(10)

where $c(M, \nu^{(-1)}) = (\pi^{(M-3)/2}) \Gamma((\nu^{(-1)} - M)/2))^{-1}$. By combining the Gaussian likelihood with unknown precision parameter $a^{(i)}$ and a Wishart distribution of $a^{(i)}$, a posterior distribution is produced as a new Wishart distribution with the updated hyperparameters. A reproducible prior/posterior pair of Wishart distributions is formulated for incremental adaptation. After the ICA parameters $\theta^{(i)}$ are estimated by the first stage using (5), we perform the transformation-based adaptation

$$G_{\eta^{(i)}}(a^{(i)}) = \eta^{(i)} a^{(i)}$$

(11)

where the transformation matrix $\eta^{(i)}$ is estimated by the second stage using (6). The dynamic source separation is executed by alternatively and iteratively estimating ICA parameters $\theta^{(i)}$ and compensation parameters $\eta^{(i)}$. The newest mixing coefficients and the number of sources can be tracked in NB-ICA procedure.

3.3 Model inference

Since the exact posterior distribution $p(\theta | X_i)$ does not exist, we employ VB inference [9] to approximate the posterior by a variational distribution $q(\theta)$. In VB framework, we maximize the lower bound of the logarithm of marginal distribution or equivalently minimize the Kullback-Leibler distance between true posterior and approximate posterior. Considering the factorized variational inference

$$q(\theta^{(i)}) = q(\theta^{(i)}) q(s^{(i)} | a^{(i)} q(\theta^{(i)})$$

(12)

, the lower bound or the negative free energy is obtained by

$$\log p(X_i) \geq \log p(X_i | A^{(i)}, s^{(i)}, a^{(i)}, \theta^{(i)}) > q_{\theta^{(i)}, \theta^{(i)}} + S(q_{\theta^{(i)})} + <\log p(A^{(i)} | a^{(i)}) > q_{\nu^{(i)}}, S(q_{\nu^{(i)}}, S(q_{\nu^{(i)}}, q_{\nu^{(i))} > q_{\nu^{(i))} + S(q_{\nu^{(i))} + <\log p(\theta^{(i)}) > q_{\nu^{(i))} + <\log p(a^{(i)}) > q_{\nu^{(i))}$$

(13)

where $S(q)$ is the entropy of a distribution $q()$. By taking the differential of lower bound with respect to $q(\theta^{(i)})$, we can find the variational parameters in $q(\theta^{(i)})$, $q(s^{(i)} | a^{(i)})$ and $q(\theta^{(i)})$. Similar solution can be found in [1]. The mixing matrix and the source signals are consequently estimated. The hyperparameters $\phi^{(-1)} = \{\nu^{(-1)}, A^{(-1)}\}$ are updated to $\phi^{(i)} = \nu^{(i)}, \eta^{(i)}$. In this study, the converged variational parameters at frame $t$ are propagated as the initial variational parameters at frame $t+1$. The precision of mixing matrix $a^{(i)}$ is updated incrementally.

4. EXPERIMENTS

4.1 Experimental setup

In the experiments, we evaluated the proposed NB-ICA for dynamic source separation. We simulated the scenarios with dynamic mixing matrices and dynamic sources by using the speech signals in ICA’99 Test Sets (http://sound.media.mit.edu/ica-bench/). Figure 2 (a) shows the dynamic source signals with two speakers. The scenario of source switching and moving was considered. The scenario of presence and absence of the same speaker was examined. The nonstationary mixing matrix was given by $A^{(i)} = [(\cos(2\pi f_1t) - \sin(2\pi f_1t))(\sin(2\pi f_2t) \cos(2\pi f_2t))]$ where $f_1$ and $f_2$ denote the changing rate of mixing coefficients. The mixing matrix represents the relative position and distance between sources and sensors. The first and the second sources to the sensors had a changing rate of $f_1 = 1/5$ Hz and $f_2 = 1/2.5$ Hz, respectively. We implemented the online signal separation with a time window of 0.25 sec. For comparison, the variational Bayesian ICA (VB-ICA) [1], switching ICA [7], Bayesian ICA with hidden Markov sources (BICA-HMM) [5] and online VB-ICA [8] were implemented. The demixed signals of NB-ICA were displayed in the Figure 2 (a). The signal-to-interference ratios (SIRs) of two demixed signals were reported in Table 1. NB-ICA obtained the highest SIR among different methods. In our investigation, NB-ICA significantly saves 97% of running time.

Figure 1: Graphical representation for NB-ICA model.
compared to the switching ICA when implementing on a personal computer with a core 2 quad 2.4GHz CPU and a memory of 2 GB.

Figure 2: (a) Waveforms of source signals, mixed signals and demixed signals using NB-ICA. (b) Comparison of negative free energy with (pink) and without (green) compensation scheme.

4.2 Evaluation of compensation scheme
Next, we investigate the effect of compensation scheme of (10) in NB-ICA method. Figure 2 (b) shows the negative free energy calculated by using the same demixed signals as displayed in Figure 2 (a). The results with and without applying compensation parameter $a^{(t)}$ at each time window $t$ were compared. Higher negative free energy implies that better matching score is measured from the demixed signals. It is obvious that compensation scheme did help tracing the variations of ARD parameters. This scheme provides a robust parameter guess for the next time window. Also, we look at the source signals during the period between 3 sec and 8 sec. Figure 3 shows the waveforms of source signals and the corresponding curves of ARD parameters $a^{(t)}$. The estimated ARD reflected the number of latent sources at different time stamps. The irrelevant sources could be pruned by considering this ARD value. ARD did effectively indicate the activities of source signals in time series. During the period between 5.5 sec and 7 sec, the first source had a silence segment and was clearly reflected by ARD. Using this method, we don’t need to use additional variable to represent the status of source signals. The status of each source signals was detected from the mixed observations.

Table 1: Comparison of SIRs for different methods

<table>
<thead>
<tr>
<th></th>
<th>VB-ICA</th>
<th>Switching ICA</th>
<th>BICA-HMM</th>
<th>Online VB-ICA</th>
<th>NB-ICA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demixed signal 1</td>
<td>-1.4367</td>
<td>-2.3636</td>
<td>-1.968</td>
<td>3.3426</td>
<td>9.3373</td>
</tr>
<tr>
<td>Demixed signal 2</td>
<td>-1.9844</td>
<td>-1.1152</td>
<td>-1.438</td>
<td>5.1722</td>
<td>10.329</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS
This paper presented a novel NB-ICA algorithm to cope with the problem of dynamic source separation. The online VB learning adaptively captured the statistics of the source signals and the status of each source. This learning scheme was run in a real-time manner. The estimated mixing matrix served as a robust initial value in VB learning at the next time frame since the variations of mixing matrix was compensated. Instead of incorporating the Markov model for individual source signals using batch estimation, the proposed NB-ICA employed an ARD model and efficiently determined the number of the latent sources by incremental learning method. Without incorporating the Markov model in the Bayesian ICA framework, the computational cost was alleviated. This method was feasible for practical signal separation.

6. REFERENCES