University Of Waterloo

CS 686

Introduction to Artificial Intelligence

Bayesian Approach for Blind Source Separation

Author:
Heming Wang

Supervisor:
Prof. Richard Mann &
Prof. Edward Vrscay

April 7, 2017

1 Introduction

Blind source separation(BSS) is an important research topic in signal processing and machine learning, and It is widely applied in various areas like audio and image processing. This topic comes from the famous cocktail party problem, which states that people can focus on a single party speaker while there are other conversations and disturbance. In academic, this problem is actually filter out single speech source from a mixture of speeches and noise. The problem addressed in BSS is that when the number of sensors is smaller than the number of signal sources, how to separate unobservable signal sources from signal mixtures[1].

In general, the source separation problem can be formulated by the following equation

$$x_i = \sum_{j=1}^{M} a_{ij} s_j \tag{1}$$

Where M represents the number of sources, and a_{ij} is the amplitude parameter for the sources s_j. x_i is the signal mixtures. The sources are assumed to be independent of each other. When the number of sensors is less than the number of sources, we usually do not have enough information to accurately recover the sources. Therefore, we need some strong prior knowledge about the mixtures, and then it will be possible for us to find a statistically optimal separation that has the greatest probability. Using the Gaussian prior, we can keep the inference tractable so that Bayesian estimator can be applied for the source mixtures.

In this project, my implementation will follow the Bayesian statistical approach described in this paper[2]. The audio mixture in modeled in the transformed field(time-frequency field) based on Gaussian Mixture Model. And the signals mixtures is expressed as the linear sum of independent sources plus noise.

Suppose the mixture has two signal components s_1 and s_2. In general, we can estimate the signal x through probabilistic approach by finding the maximum likelihood

$$(\hat{s_1}, \hat{s_2}) = \underset{s_1, s_2}{\operatorname{argmax}}\, p(x|s_1, s_2) \tag{2}$$

Using Bayesian equation, we can obtain that

$$p(s_1, s_2) \propto p(x|s1, s2) p_1(s_1) p_2(s_2) \tag{3}$$

Here we use the assumption that the sources are independent to each other, i.e. $p(s_1, s_2) = p_1(s_1) \cdot p_2(s_2)$. The equation (3) is the basis of our estimation. Given priori information of the sources, it is possible to obtain optimal parameters by training on signal data.

2 Gaussian Mixture Model

In this section, we will introduce the basics about Gaussian Mixture model and discuss the techniques that parameterizing audio spectrum based on the GMM. The normal approach is to use the Monte Carol method; however, this method is computationally expensive and slow when estimating the parameters. In this project, I apply the method mentioned in [3], which does not require a lot of computation, and estimating result is as expected.

2.1 Introduction for Gaussian Mixture Models

The Bayesian formalism offers a natural framework for us the utilize the prior knowledge in the estimation problem. The Gaussian Mixture Model assumes that the signal is consist of M Gaussian components. In other words, each data point belongs to one of the components and we need to infer the Gaussian distribution of each component. To represent it mathematically, we formulate the model in terms of latent variables z, which corresponds to the mixture component. In general, the mixture model assumes that we first sample z, and the sample the observables x from the Gaussian distribution that depends on z[4].

$$p(z, \mathbf{x}) = p(z)p(\mathbf{x}|z) \tag{4}$$

Such a model is referred as a mixture of Gaussians(MoG).

2.2 GMM Representation of the spectrum

In the paper of [5], it estimates the Gaussian parameters directly on the DFT magnitude spectrum.The GMM spectral estimation used the single impulse function to form the histogram. Each frequency bin is represented by an impulse function which is weighted by the normalized FFT magnitude. It applies the Monte-Carlo approach, which generates sufficiently large number of random number according to the distribution of histogram, and then use the standard form of GMM-EM algorithm to estimate the GMM parameters.

In practice, this way is computationally expensive the number of data to process is large.

The method I applied used the continuous bin probability functions to form the histogram. Each frequency bin is centered in its bar of the histogram with the width of 1.

The continuous probability density function is then formed by

$$p(x|\mathbf{g} = \sum_{i=1}^{N} P(g_i)p(x|g_i) \tag{5}$$

where $P(g_i)$ is the prior probability for the ith frequency bin g_i, and

$$p(x|g_i) = \omega \tag{6}$$

where the prior probabilities are obtained by normalizing the spectrum, i.e. $P(g_i) = \bar{s_i}(t)$.

We use the Kullbeck Leibler(KL) divergence to calculate the distance between two PDFs GMM $p(x|\theta)$ and the spectral histogram $p(x|\mathbf{g})$:

$$D(p(x|g), p(x|\theta)) = \int_{-\infty}^{+\infty} log(\frac{p(x|g)}{p(x|\theta)})dx \tag{7}$$

Where the second term can be simplified as

$$\begin{aligned} \int_{-\infty}^{+\infty} P(x|g_i)logp(x|\theta)dx &= \int_{f_i-\frac{1}{2}}^{f_i+\frac{1}{2}} P(x|g_i)logp(x|\theta)dx \\ &= Elogp(x|\theta)|g \end{aligned} \tag{8}$$

To estimate the GMM parameters, the Monte-Carlo approach draw large number(D) of data points from the histogram. In this case, the auxiliary function $Q(\theta, \hat{\theta})$ would be:

$$\begin{aligned} Q(\theta, \hat{\theta}) = \frac{1}{D}\sum_{m=1}^{M}[\sum_{d=1}^{D} P(\omega_m|x_d, \theta)log(p(x_d|\omega_m, \hat{\theta_m}))] \\ + \frac{1}{D}\sum_{m=1}^{M}[\sum_{d=1}^{D} P(\omega_m|x_d, \theta)log(\hat{P}(\omega_m))] \end{aligned} \tag{9}$$

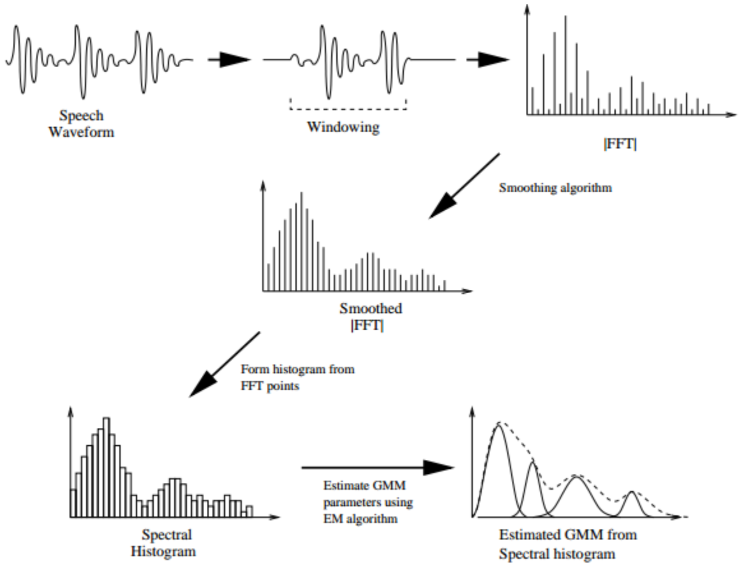

Figure 1: This is a figure caption.

To avoid large amount of computation, we make assumption that all points that are drawn will share the same posterior probability:

$$P(\omega_m|x_d,\theta) \approx P(\omega_m|g_i,\theta) \tag{10}$$

And when D approaches infinity, the prior probabilities can be considered to cone from a given bin $P(g_i)$, and the auxiliary function becomes

$$\begin{aligned} Q(\theta,\hat{\theta}) \approx \sum_{m=1}^{M}[\sum_{i=1}^{N} P(g_i) Elogp(x|\omega_m,\hat{\theta_m})|g_i] \\ +\sum_{m=1}^{M}[\sum_{i=1}^{N} P(\omega_m|g_i,\theta)P(g_i)log(\hat{P}(\omega_m))] \end{aligned} \tag{11}$$

Which can be regarded as our Bayesian estimator. And by taking the derivative of the auxiliary function (11) and set it to zero, we can obtain the parameter update

$$\hat{\mu_j} = \frac{\sum_{i=1}^{N} P(g_i)P(\omega_j|g_i,\theta)f_i}{\sum_{i=1}^{N} P(g_i)P(\omega_j|g_i,\theta)} \tag{12}$$

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{N} P(g_i)P(\omega_j|g_i,\theta)[(f_i-\hat{\mu}_j)+\frac{1}{12}]}{\sum_{i=1}^{N} P(g_i)P(\omega_j)P(\omega_j|g_i,\theta)} \tag{13}$$

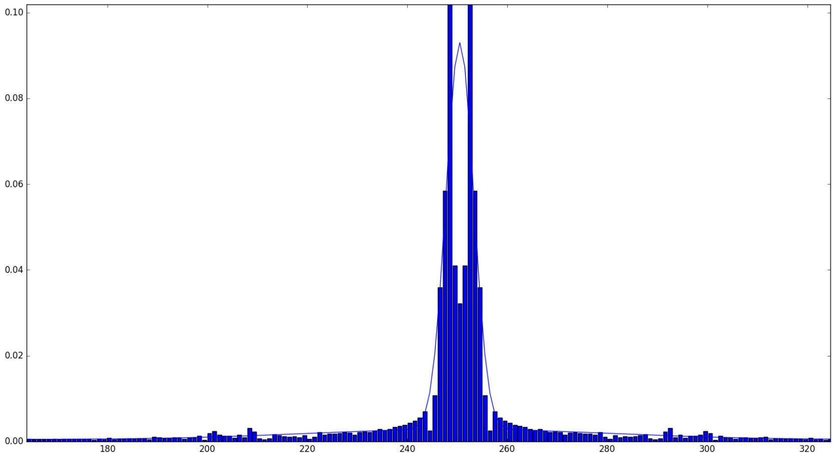

Figure 2: GMM spectrum parameter estimation

$$\begin{aligned} \hat{P}(\omega_j) &= \frac{\sum_{i=1}^{N} P(g_i)P(\omega_j|g_j,\theta)}{\sum_{m=1}^{M}\sum_{i=1}^{N} P(g_i)P(\omega_m|g_i,\theta)} \\ &= \sum_{i=1}^{N} P(g_i)P(\omega_j|g_j,\theta) \end{aligned} \tag{14}$$

Apply the EM algorithm and the parameters will converge after reasonable number of iterations. The Figure1 demonstrates the general process of GMM spectrum fitting.

3 Blind Source Separation

3.1 Transformation Domain

In order to make different sources to be independent statistically, it is crucial to choose an appropriate transform to apply. Discrete cosine transform is chosen in this project for its properties: it is unitary, robust to noise and most importantly, the transformed signal is real, which is convenient for model estimation.

3.2 Bayesian Analysis for Gaussian Mixture Model

As described in the introduction part, the signal mixtures are consist of several principle components. In this project, several methods that rely on the Bayesian approach are discussed. Signals can be expressed as $x = As + b$, where b is the noise that follows the Gaussian distribution. Generally, we assume the prior knowledge of Gaussian Mixture densities. Assume there are K Gaussian components with the proportion of π^i. In the Bayesian formalism, the prior densities are

$$p(s_1) = \sum_{i=1}^{K_1} \omega_i^1 \frac{exp[-\frac{1}{2}s_1^T \Sigma_1^{i-1} s_1]}{(2\pi)^(N/2)|det(\Sigma_1^i)|^{1/2}} \tag{15}$$

$$p(s_2) = \sum_{i=1}^{K_2} \omega_i^2 \frac{exp[-\frac{1}{2}s_2^T \Sigma_2^{i-1} s_2]}{(2\pi)^(N/2)|det(\Sigma_2^i)|^{1/2}} \tag{16}$$

where $\sum$ represents the covariance matrix, and the probability distribution should satisfy $\sum_{i=1}^{K} \omega^{(i)} = 1$.

In order to make the Bayesian estimator tractable, we introduce two hidden variables q_1 and q_2 which are associated with both signal sources, which is a typical incomplete data setting. For the hidden process, the following likelihoods are derived

$$p(s_i|q_i = k) = \frac{exp[-\frac{1}{2}s_i^T]\Sigma_i^{k-1}s_i}{(2\pi)^{N/2}|det(\Sigma_i^k)|^{1/2}} \tag{17}$$

$$p(q_i = k) = \omega_i^k \tag{18}$$

The next step is to estimate the hidden states (q_1, q_2) as they are unknown. When $q_1 = i$, the signal s_1 will have a Gaussian distribution conditionally to q_1, then the signal mixture will have Gaussian distribution conditionally to (q_1, q_2) with covariance matrix $\sum_1^i + \sum_2^j + \sigma^2 I$. Thus the a posteriori low for components (i, j) can be derived as

$$\begin{aligned} p(i,j|x) &\propto p(x|i,j) * p(i) * p(j) \\ &\propto \omega_1^i \omega_2^j Gaussian(x, \Sigma_1^i + \Sigma_2^j + \sigma^2 I) \end{aligned} \tag{19}$$

The MAP estimator for signals s_1, s_2 can be deduced from the above formula

$$\begin{aligned} -2logp(s_1, s_2|x, i, j) &= \frac{1}{\sigma_2}||x - s_1 - s_2||_2^2 \\ &= +s_1^T[\Sigma_1^i]^-1s_1 + s_2^T[\Sigma_2^j]^-1s_2 + cte. \end{aligned} \tag{20}$$

And if i and j are known, posterior mean estimators for both signals are

$$E(s_1|i, j) = \Sigma_1^i[\Sigma_1^i + \Sigma_2^j + \sigma^2 I]^-1x \tag{21}$$

$$E(s_2|i, j) = \Sigma_2^i[\Sigma_1^i + \Sigma_2^j + \sigma^2 I]^-1x \tag{22}$$

From Bayes law

$$\begin{aligned} p(s_1|x) &\propto \sum_{i,j} p(s_1|x, i, j)p(q_1 = i, q_2 = j|x) \\ &\propto \sum_{i,j} p(s_1|x, i, j)\gamma_{i,j}(x) \end{aligned} \tag{23}$$

So finally, it is natural that

$$E(s_1|x) = \sum_{i=1}^{K_1}\sum_{j=1}^{K_2} \gamma_{i,j}(x) \cdot \Sigma_1^i[\Sigma_1^i + \Sigma_2^i + \sigma^2 I]^{-1} \cdot x \tag{24}$$

Similarly for s_2.

3.3 Separation Algorithm

When the signal mixtures are transformed (such as short-term Fourier transform(STFT)). At each time index t, we calculate the weighting parameter, and then we calculate the posterior mean estimator derived before to estimate the signals.

The algorithm is tested on a speech mixtures made by software audacity. The speech mixtures consist of a male speaker piece and a female speaker piece.

The separation result is evaluated by the Source to Artifact Ratio(SAR). and a SAR is calculated by

$$SAR = 20log_{10}\frac{s}{n} \tag{25}$$

Algorithm for Blind Source Separation
1. Firstly apply STFT to the signal mixtures
2. GMM parameter estimation on the spectrum
4. Then we set initialization parameters for two sources
5. For every time frame t, calculate the weighting probabilities $\gamma_{i,j}$
6. Finally we calculate the posterior mean estimator in the time-frequency domain
7. Recover the signals

Table 1: Algorithm table.

In my experiment, I calculated the SAR and the result is -2.40dB for male speaker and 2.40dB for female speaker signal , which is a reasonable result compared with the decibel tables in the paper [2].

4 Further research potential

This project is working on the Short time Fourier Transform, other transforms may have better performance since the separation is on the magnitude of frequency spectrum and the phase information is ignored in the separation process. Transforms like Discrete Cosine Transform, Hilbert Transform, Empirical Mode Decomposition(EMD) can be applied to see the separation performance. In addition, different Bayesian method MAP, ML and Posterior Mean (PM) can be applied together and compared.

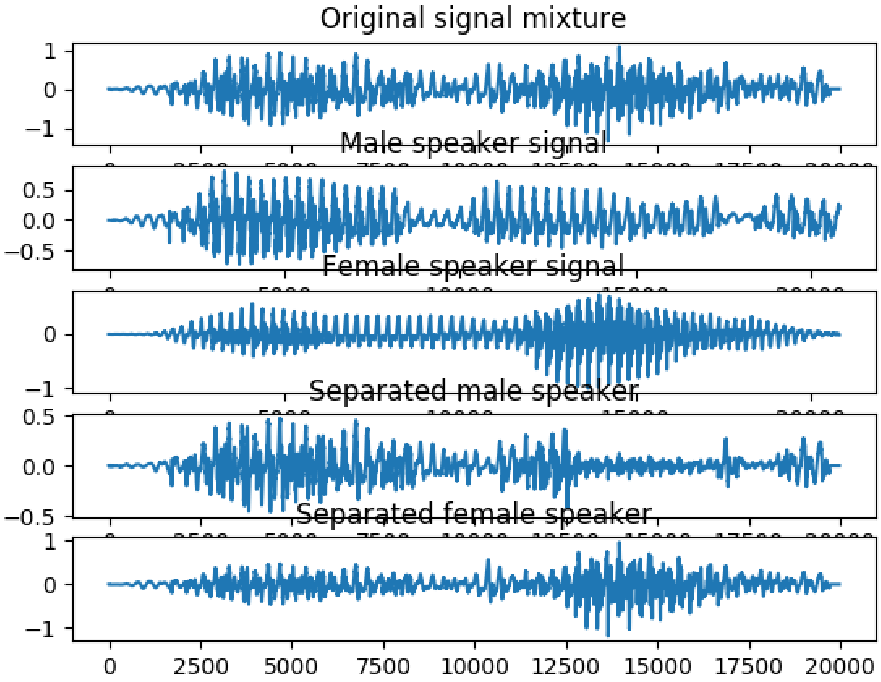

Figure 3: Separation result for speech mixtures made by Audacity. First graph is the mixture of two speech signals(a male and female speaker), second is the original signal1(a male speaker), the third graph is the original signal 2(a female speaker). The last two figures are the recovered signals using our Bayesian source separation approach, which apparently has a lot of noise, but of the same wave form as the original one

References

[1] B. A. Pearlmutter, L. C. Parra *et al.*, "Maximum likelihood blind source separation: A context-sensitive generalization of ica," *Advances in neural information processing systems*, pp. 613–619, 1997.

[2] L. Benaroya, F. Bimbot, and R. Gribonval, "Audio source separation with a single sensor," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 191–199, 2006.

[3] M. N. Stuttle, "A gaussian mixture model spectral representation for speech recognition," Ph.D. dissertation, University of Cambridge, 2003.

[4] D. Reynolds, "Gaussian mixture models," *Encyclopedia of biometrics*, pp. 827–832, 2015.

[5] P. Zolfaghari and T. Robinson, "Formant analysis using mixtures of gaussians," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 2. IEEE, 1996, pp. 1229–1232.