ICA를 위한 Generalized 가우시안 Prior

최승진
충북대학교 전기공학과
schoi@engine.chungbuk.ac.kr

GENERALIZED GAUSSIAN PRIOR FOR ICA

Seungjin CHOI
Department of Electrical Engineering, Chungbuk National University, KOREA

요 약

Independent componet analysis (ICA)는 주어진 데이터를 통계적으로 독립인 요소들의 선형 결합으로 표시하는 통계학적 방법이다. ICA의 주요한 적용 분야중의 하나는 source들의 선형 mixture로부터 어떠한 서전 정보도 없는 상태에서 원래의 통계학적 독립변수인 source를 복원하는 blind separation이다. ICA와 sourece separation을 위한 다양한 신경 학습 알고리듬이 제시되어왔다. ICA의 학습 알고리듬에서는 비선형 함수가 중요한 역할을 한다. 이 논문에서는 generalized 가우시안 prior를 도입하여 다양한 확률분포를 갖는 source들의 mixture를 분리하는 효율적인 source separation알고리듬을 제시한다. 모의실험을 통하여 제안된 방법의 우수성을 살펴본다.

1. INTRODUCTION

ICA has recently received extensive attention because of its applications in source separation and feature extraction. The goal of ICA is to ecomposemultivariate data into a linear sum of statistically independent components. ICA is a widely-used approach to source separation, the goal of which is torecover unknown sources (latent variables) from their linear instantaneous mixtures, without resorting to any prior information except for the assumption of statistical independence of sources. In the context of source separation, it is assumed that a vector of n sensor signals denoted by $x(t) = [x_1(t) \cdots x_n(t)]^T$ has the form

$$x(t) = As(t), \tag{1}$$

where $A \in R^{n \times n}$ is called *mixing matrix* and $s(t)$ is an n-dimensional vector whose elements are called *sources*.

Source separation aims at designing a demixing filter W such that the filter output vector $y(t) = Wx(t)$ is a possibly rescaled and reordered source vector [6,2]. In other words, it is required to transform the observation vector $x(t)$ into $y(t)$ such that the following decomposition is satisfied:

$$y(t) = P\Lambda s(t) \tag{2}$$

where P is the permutation matrix and Λ is some nonsingular diagonal matrix.

Most learning algorithms [5, 1, 4] for source separation have the form

$$W(t+1) = W(t) + \eta_t F(W(t), y(t)), \tag{3}$$

where $F(W(t), y(t))$ is a estimating function. Examples of estimation functions are

$$F(W(t), y(t)) = \{I - \varphi(y(t))y^T\} W(t), \tag{4}$$

and

$$F(W(t), y(t)) = \{I - y(t)y^T(t) - \varphi(y(t))y^T(t) + y(t)\varphi^T(y(t))\} W(t). \tag{5}$$

The parameter η_t represents a learning rate that is properly specified in advance. The nonlinear function $\varphi(\cdot)$ is given by

$$\varphi(y) = [\varphi_1(y_1) \cdots \varphi_n(y_n)]^T, \tag{6}$$

where the time index has been dropped for the sake of simplicity.

The optimal form of the nonlinear function $\varphi_i(y_i)$ is the negative score function which is given by

$$\varphi_i(y_i) = -\frac{d \log p_i(y_i)}{dy_i}, \tag{7}$$

where $p_i(\cdot)$ is the probability density function of the source s_i, which is called *prior*.

In order to select nonlinear functions $\{\varphi_i(\cdot)\}$, we need to have the good estimates of source distributions. However, in practice, it is difficult to estimate the probability distributions of sources, since in fact we do not what type of sources are in there.

In this paper we present a efficient method to learn the form of nonlinear function in order to achieve successful source separation. The approach we take here is to use a parameterized density model which can approximated most uni-modal distributions.

2. GENERALIZED GAUSSIAN DENSITY MODEL

2.1. Generalized Gaussian Distribution

The *generalized gaussian* probability distribution (see Figure 1 is a set of distributions parameterized by a positive real number α, which is usually referred to as the *Gaussian exponent* of the distribution. The Gaussian exponent α controls the "peakiness" of the distribution. The probability density function (PDF) for a generalized Gaussian is described by

$$p(y;\alpha)=\frac{\alpha}{2\lambda\Gamma(1/\alpha)}e^{-|y/\lambda|^{\alpha}} \tag{8}$$

where $\Gamma(x)$ is Gamma function. Note that if $\alpha=1$, the distribution becomes the standard "Laplacian" distribution. For $\alpha=2$, the distribution is standard normal distribution, and when α approaches ∞ the distribution becomes uniform.

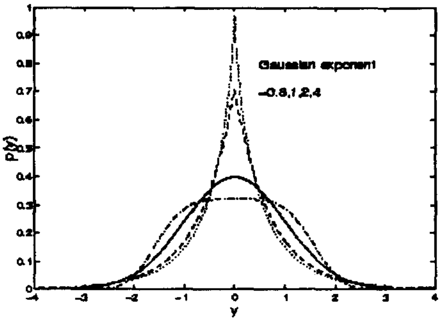

Figure 1: The generalized Gaussian distribution is plotted for several different values of gaussian exponent, α= 0.8, 1, 2, 4.

2.2. The Moments of the Generalized Gaussian Distribution

In order to fully understand the properties of the generalized Gaussian distribution, it is useful to look at its moments (specially 2nd and 4th moments which give the kurtosis).

The 2nd moment of the generalized Gaussian distribution is computed as

$$m_2=\lambda^2\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}, \tag{9}$$

and the 4th moment is

$$m_4=\lambda^4\frac{\Gamma(5/\alpha)}{\Gamma(1/\alpha)}. \tag{10}$$

2.3. Kurtosis and Gaussian Exponent

The kurtosis is a nondimensional quantity. It measures the relative peakedness or flatness of a distribution. A distribution with positive kurtosis is termed *leptokurtic* (super-Gaussian). A distribution with negative kurtosis is termed *platykurtic*(sub-Gaussian). The kurtosis of the distribution is defined in terms of the 2nd- and 4th-order moments as

$$\kappa(y)=\frac{m_4}{m_2^2}-3, \tag{11}$$

where the constant term -3 makes the value zero for standard normal distribution.

For a generalized Gaussian distribution, the kurtosis can be expressed in terms of the Gaussian exponent, given by

$$\kappa_\alpha=\frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}-3. \tag{12}$$

The plot of kurtosis κ_α versus the Gaussian exponent $\alpha$for leptokurtic and platykurtic signals are shown in Figure 2.

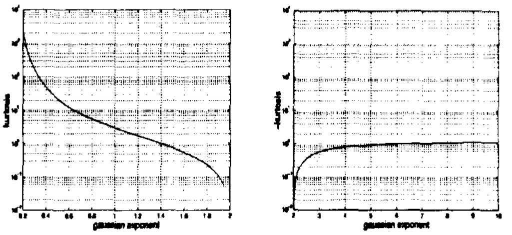

Figure 2: The plot of kurtosis κ_α versus Gaussian exponent α: (a) for leptokurtic signal; (b) for platykurtic signal.

3. THE ALGORITHM

The generalized Gaussian density function is given by i.e.

$$p_i(y_i)=\frac{\alpha}{2\lambda\Gamma(1/\alpha_i)}e^{-|y_i/\lambda|^{\alpha_i}} \tag{13}$$

With the density model in (13), the (normalized) nonlinear function $\varphi_i(y_i)$ is calculated as

$$\begin{aligned}\varphi_i(y_i)&=\frac{d\log p_i(y_i)}{dy_i}\\&=|y_i|^{\alpha-1}sgn(y_i)\end{aligned} \tag{14}$$

where $sgn(y_i)$ is the signum function of y_i. Note that for $\alpha=1$, (14) becomes a signum function (where can also be derived from the Laplacian density medel), for $\alpha=4$, (14) becomes a cubic function which is known to be a good choice for sub-Gaussian source. In order to select a proper value of Gaussian exponent α_i, we estimate the kurtosis of the output signal y_i and select the corresponding α_i from Figure 2. The kurtosis of y_i, κ_i can be estimated via the following iterative algorithm:

$$\kappa_i(t+1)=\frac{m_4(t+1)}{m_2^2(t+1)}-3, \tag{15}$$

where

$$m_{4i}(t+1)=(1-\delta)m_{4i}(t)+\delta|y_i(t)|^4, \tag{16}$$

$$m_{2i}(t+1)=(1-\delta)m_{2i}(t)+\delta|y_i(t)|^2, \tag{17}$$

where δ is a small constant, say, 0.01.

In general, the estimated kurtosis of y_i that is calculated in (15) does not exactly correspond the kurtosis of source s_i. However, at least we can decide whether y_i belongs to either sub-Gaussian signal. Cardoso [3] pointed out that the reasonable mismatch between the hypothesized density and the true density does not degrade the performance of the source separation algorithms (4) and (5). for sub-Gaussian source we observe that the kurtosis does not change much with respect to the Gaussian exponent (see Figure 2). Thus in practice, we set $\alpha_i=4$ as long as the estimated kurtosis of y_i is negative. For super-Gaussian source, it might be enough to use two or three different values of α_i depending what range the estimated kurtosis falls in. for example we suggest $\alpha_i = 0.8, \alpha_i = 1$, and $\alpha_i = 1.3$.

4. SIMULATIONS

We have performed simulations with eight different sources (see Figure 3). Sources consist of digitized speech, the sound of a bell, laughing, music, chirping, a steam-whistle, and two noises (one has uniform distribution and the other has Gaussian distribution).

Mixture signals were generated by the mixing matrix $\boldsymbol{A}$ whose elements were drawn from uniformly distributed random numbers in [0,1]. Mixture signals are shown in Figure 4. One can see that all the observations are close to Gaussian distributions. It is easily justified by the central limit theorem.

In the simulation, we have applied the algorithm (5) with our method of selecting nonlinear functions. The learning rate $\eta_t = .001$ was used and the demixing filter $\boldsymbol{W}$ was initially set as the identity matrix. The recovered signals and their corresponding spectrograms are shown in Figure 5. One can observe that sources are almost perfectly recovered and densities are matching to their true ones.

5. CONCLUSIONS

We have introduced the generalized Gaussian density model that is able to approximate a wide class of super- and sub-Gaussian distributions, in the framework of natural gradient based ICA algorithms. From the generalized Gaussian density model, we have derived a new self-adaptive nonlinear function which is controlled by only one parameter(Gaussian exponent). We have also introduced a simple practical implementation where several different values of Gaussian exponent was chosen properly according to the estimated kurtosis of demixing filter output. Stability for several different choices of α was analyzed. We have also considered both on-line and batch versions of flexible ICA algorithms. Useful behavior of the proposed approach was demonstrated by computer simulation results.

6. ACKNOWLEDGMENTS

The suthor wish to thank Profs S. Amari and A. Cichocki for their helpful discussions in Brain Science Institute, RIKEN, JAPAN. This work was supported by Braintech 21, Ministry of Science and Technology.

7. REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In D. D. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Nueral Infotmation Processing Systems*, volume 8, pages 757-763. MIT press, 1996.

[2] A. Bell and T. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129-1159, 1995.

[3] J. F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Signal Processing Letters*, 4:112-114, 1997.

[4] J. F. Cardoso. and B. H. Laheld. Equivariant adaptive source separation. *IEEE Trans. Signal Processing* 44:3017-3030, 1996.

[5] A. Cichocki, R. Unbehauen, and E. Rummert. Robust learning algorithm for blind separation of signals. *Electronics Letters*, 43(17):1386-1387, 1994.

[6] P. Comon. Independent component analysis, a new concept? *Signal Processing* 36(3):287-314, 1994.

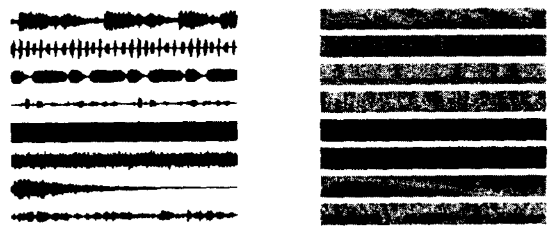

Figure 3: Original sound sources and their corresponding spectrograms.

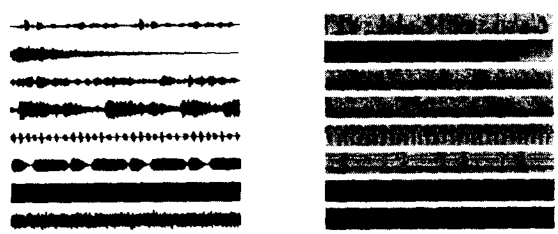

Figure 4: mixtures and their corresponding spectrograms.

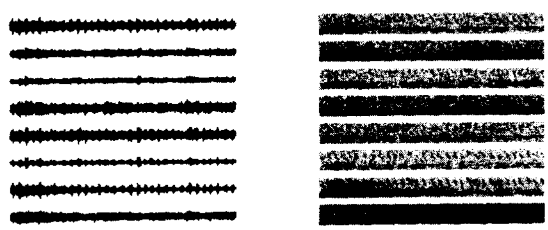

Figure 5: Recovered signals using the proposed algorithm and their corresponding specrtograms.