Copula Component Analysis

Jian Ma and Zengqi Sun

Department of Computer Science,
Tsinghua University, Beijing 100084, China
majian03@mails.tsinghua.edu.cn

Abstract. A framework named copula component analysis (CCA) for blind source separation is proposed as a generalization of independent component analysis (ICA). It differs from ICA which assumes independence of sources that the underlying components may be dependent by certain structure which is represented by Copula. By incorporating dependency structure, much accurate estimation can be made in principle in the case that the assumption of independence is invalidated. A two phrase inference method is introduced for CCA which is based on the notion of multi-dimensional ICA. Simulation experiments preliminarily show that CCA can recover dependency structure within components while ICA does not.

1 Introduction

Blind source separation (BSS) is to recover the underlying components from their mixtures, where the mixing matrix and distribution of the components are unknown. To solve this problem, independent component analysis (ICA) is the most popular method to extract those components under the assumption of statistically independence[1,2,3,4,5]. However, in practice, the independence assumption of ICA cannot always be fully satisfied and thus strongly confines its applications. Many works have been contributed to generalize the ICA model,[6] such as Tree-ICA[7], Topology ICA[8]. A central problem of those works is how to relax the independent assumption and to incorporate different kinds of dependency structure into the model.

Copula [9] is a recently developed mathematical theory for multivariate probability analysis. It separates joint probability distribution function into the product of marginal distributions and a Copula function which represents the dependency structure of random variables. According to *Sklar theorem*, given a joint distribution with margins, there exists a copula uniquely determined. Through Copula, we can clearly represent the dependent relation of variables and analysis multivariate distribution of the underlying components.

The aim of this paper is to use Copula to model the dependent relations between elements of random vectors. By doing this, we transform BSS into a parametric or semi-parametric estimation problem which mainly concentrate on the estimation of dependency structure besides identification of the underlying components as ICA do.

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 73–80, 2007.

This paper is organized as follows: we briefly review ICA and its extensions in section 2. The main conclusions of copula theory are briefly introduced in section 3. In section 4, we propose a new model for BSS, named copula component analysis (CCA) which takes dependency among components into consideration. Inference method for CCA is presented in section 5. Simulation experiments are presented in section 6. Finally, we conclude the paper and give some further research directions.

2 ICA and Its Extensions

Given a random vector $\mathbf{x}$, ICA is modeled as

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where the source signals $\mathbf{s} = \{s_1, \ldots, s_n\}$ assume to be mutually independent, $\mathbf{A}$ and $\mathbf{W} = \mathbf{A}^{-}$ is the invertible mixing and demixing matrix to be solved so that the recovered underlying components $\{s_1, \ldots, s_n\}$ is estimated as statistically independent as possible.

Statistical independence of sources means that the joint probability density of $\mathbf{x}$ and $\mathbf{s}$ can be factorized as

$$\begin{aligned} p(\mathbf{x}) = p(\mathbf{A}\mathbf{s}) =\mid det(\mathbf{W}) \mid p(\mathbf{s}) \\ p(\mathbf{s}) = \textstyle\prod_{i=1}^{n} p_i(s_i) \end{aligned} \tag{2}$$

The community has presented many extensions of ICA with different types of dependency structures. For example, Bach and Jordan [7] assumed that dependency can be modeled as a tree (or a forest). After the contrast function is extended with T-mutual information, Tree-ICA tries to find both a mixing matrix A and a tree structure T by embedding a Chow-Liu algorithm into algorithm. Hyvärinen etc. [8] introduced the variance into ICA model so as to model dependency structure. Cardoso generalized the notion of ICA into multidimensional ICA using geometrical structure.[6]

3 A Brief Introduction on Copula

Copula is a recently developed theory which separates the margin law and the joint law and therefore gives dependency structure as a function. According to Nelson [9], it is defined as follows:

Definition 1 (Copula). *A bidimensional copula is a function $C(x, y) : I^2 \mapsto I$ with following properties:*

1. $(x, y) \subset I^2$
2. $C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \geq 0$*, for* $x_1 \leq x_2$ *and* $y_1 \leq y_2$*;*
3. $C(x, 1) = x$ *and* $C(1, y) = y$.

It's not hard to know that such defined $C(x, y)$ is a cdf on I^2. Multidimensional version can be generalized in a same manner which presents in [9].

Theorem 1 (Sklar Theorem). *Given a multidimensional random vector* $\mathbf{x} = (x_1, \ldots, x_n) \in R^n$ *with its corresponding distribution function and density function* $u_i = F_i(x_i)$ *and* $p_i(x_i), i = 1, \ldots, n$. *Let* $F(x) : R^n \mapsto I$ *denotes the joint distribution, then there exists a copula* $C(\cdot) : I^n \mapsto I$ *so that*

$$F(\mathbf{x}) = C(\mathbf{u}). \tag{3}$$

where $\mathbf{u} = (u_1, \ldots, u_n)$.

If the copula is differentiable, the joint density function of $F(\mathbf{x})$ *is*

$$P_{1,\ldots,n}(\mathbf{x}) = \prod_{i=1}^{n} p_i(x_i) C'(\mathbf{u}). \tag{4}$$

where $C'(\mathbf{u}) = \frac{\partial C(\mathbf{u})}{\partial u_1, \ldots, \partial u_n}$.

Given a random vector $\mathbf{x} = (x_1, \ldots, x_n)$ with mutually independent variables, and their cdf $F(\mathbf{x}) = \prod_i F_i(x_i)$. It is easy to obtain that the corresponding copula function called *Product Copula* is $C(\mathbf{u}) = \prod_i u_i$ and $C'(\mathbf{u}) = 1$.

4 Copula Component Analysis

4.1 Geometry of CCA

As previously stated, ICA assumes that the underlying components are mutually independent, which can be represented as (1). CCA also use the same representation (1) as ICA, but without independence assumption. Here, Let the joint density function represents by Copula:

$$p_c(\mathbf{x}) = \prod_{i=1}^{N} p_i(x_i) C'(\mathbf{u}) \tag{5}$$

where the dependency structure is modeled by function $C(\mathbf{u})$.

The goal of estimation is to minimize the distance between the 'real' pdf of random vector $\mathbf{x}$ and its counterpart of the proposed model. Given a random vector $\mathbf{x}$ with pdf $p(\mathbf{x})$, the distance between $p(\mathbf{x})$ and $p_c(\mathbf{x})$ in a sense of *K-L* divergence can be represented as

$$\begin{aligned} D(p \| p_c) &= E_{p(\mathbf{x})} \log \frac{p(\mathbf{x})}{p_c(\mathbf{x})} \\ &= E_{p(\mathbf{x})} \log \frac{p(\mathbf{x})}{\prod_i p_i(x_i)} - E_{p(\mathbf{x})} \log C'(\mathbf{u}) \end{aligned} \tag{6}$$

The first term on the right of (6) is corresponding to the K-L divergence between $p(x)$ and ICA model and the second term is corresponding to entropy of copula $C(x)$.

Theorem 2. *Given a random vector* $\mathbf{x} = (x_1, \ldots, x_n) \in R^n$ *with pdf* $p(\mathbf{x})$ *and its joint pdf* $p_c(\mathbf{x}) = \prod_{i=1}^{n} p_i(x_i) C'(\mathbf{u})$, *where* $u_i = F_i(x_i)$ *is the cdf of* x_i *and dependency structure is presented by copula function* $C(\mathbf{u}) : I^n \mapsto I, \mathbf{u} \in R^n$ *and* $C'(\mathbf{u}) = \frac{\partial^n C(\mathbf{u})}{\partial u_1, \ldots, \partial u_n}$ *is the derivative of* $C(\mathbf{u})$. *The K-L divergence* $D(p\|p_c)$ *is as*

$$D(p\|p_c) = I(x_1, \ldots, x_n) + H(C'(\mathbf{u})). \tag{7}$$

where $H(\cdot)$ *is the Shannon differential entropy.*

That is, the K-L divergence between $p(\mathbf{x})$ and $p_c(\mathbf{x})$ equal to the sum of the mutual information $I(\mathbf{x})$ and copula entropy H for function $\mathbf{u} \sim C'$.

Using the invariant of K-L divergence, we now have the following corollary to *Theorem 2* for BSS problem $\mathbf{s} = \mathbf{W}\mathbf{x}$.

Corollary 1. *With the same denotation of Theorem 2, the K-L divergence for BSS problem is*

$$D(p\|p_c) = I(s_1, \ldots, s_n) + H(C'(\mathbf{u}_s)), \tag{8}$$

where $\mathbf{u}_s$ *denotes the marginal variable for sources* $\mathbf{s}$. *Assume that the number of sources equals to that of observations.*

In other words, the distance between ICA model and the true model is presented by dependency structure and its value equals to entropy of the underlying copula function. It can be easily learned from (7) that if dependency structure was incorporated into model, the distance between data and model can be further closer than that of ICA model.

ICA is a special case when it assumes mutual independence of underlying components. Actually, ICA only minimizes the first part of (7) under the assumption of independence. This also explains why sometime ICA model is not applicable when dependency relations between source components exist.

4.2 Multidimensional ICA

From the notion of multidimensional ICA generalized from ICA by Cardoso [6], it can be derived that

$$\begin{aligned} p(\mathbf{x}) &= \prod_{k=1}^{m} p_k(\mathbf{x}_k) = \prod_{k=1}^{m} p_k(x_{i_k}, \ldots, x_{i_{k+1}-1}) \\ &= \prod_{k=1}^{m} \prod_{l=i_k}^{i_{k+1}-1} p_k(x_l) C'_k(\mathbf{u}_k) = \prod_{i=1}^{n} p_i(x_i) \prod_{k=1}^{m} C'_k(\mathbf{u}_k) \end{aligned} \tag{9}$$

where $C_k(\cdot)$ is the copula with respect to $p_k(\cdot)$. On the other side, the definition of copula gives

$$p(\mathbf{x}) = \prod_{i=1}^{n} p_i(x_i) C'(\mathbf{u}) \tag{10}$$

According to Sklar theorem, if all $p_i(\cdot)$ exist, then $C(\cdot)$ is unique. Therefore, we can derive the following result.

Theorem 3. *The copula corresponding to multidimensional ICA is factorial if all the marginal pdf of component exist, that is*

$$C'(\mathbf{u}) = \prod_{k=1}^{m} C'_k(\mathbf{u}_k) \tag{11}$$

Proof. Because of the unique of C, the above (11) can be easily derived by comparing (9) and (10).

The theorem can guide hypothesis selection of copula. That is, Copula should be factorized as a product of sub-function with different type for dependency structure of different sub-space.

Combining (7) and (11), we can derived the following:

$$D(p\|p_c) = I(u_1, \ldots, u_n) + \sum_{k=1}^{m} H(C'_k) \tag{12}$$

It means that the distance between the true model and ICA model composes of entropy of Copulas which corresponds to every un-factorial ICs spaces. Therefore, if we want to derive a model much closer to the 'true' one than ICA, we should find dependency structure of each space, that is, approach the goal step by step. This is one of the guide principles for designing algorithm of copula component analysis.

5 Inference of CCA

5.1 General Framework

In this section, we study inference method for CCA based on the notion of multidimensional ICA. Suppose the underlying copula function parameterized by $\theta \in \Theta$, thus the estimation of CCA should infer the demixing matrix $\mathbf{W}$ and θ. According to theorem 2, estimation of the underlying sources through our model requires the minimization of the K-L divergence of (7) or (12). Thus the objective function is

$$\min D(p\|p_c; \mathbf{W}, \theta) \tag{13}$$

which composes of two sub-objective: $\min I(x_1, \ldots, x_n; \mathbf{W})$ and $\min H(C'(\mathbf{u}); \mathbf{W}, \theta)$. Because $\mathbf{u}$ in the latter objective depends on the structure of IC spaces derived from the former objective, we should handle the optimal problem min $I(x_1, \ldots, x_n; \mathbf{W})$ at first. The first objective can be achieved by ICA algorithm. For the second one we proposed the Infomax like principle given a parametric family of copula.

We propose that the framework of CCA composes of two phrases:

1. Solve $\mathbf{W}$ through minimization of mutual information .
2. Determine $\mathbf{W}$ and θ so that the objective function (13) is minimized.

5.2 Maximum Likelihood Estimation

Given the parametric model of Copula, maximum likelihood estimation can be deployed under the constraint of ICA. Consider a group of independent observations $x_1, \ldots, x_T$ of $n \times 1$ random vector $\mathbf{X}$ with a common distribution $\mathcal{P} = C'_\theta(\mathbf{x}) \prod_{i=1}^{T} p_i(x_i); \theta \in \Theta$ where $p_i(x_i)$ is marginal distribution associated with x_i, and the log-likelihood is

$$\begin{aligned}\mathcal{L}(\mathbf{W}, \theta) &= \frac{1}{T} \log C'_\theta(\mathbf{x}) \prod_{i=1}^{T} p_i(x_i) \\ &= \frac{1}{T} \sum_{i=1}^{T} \log p_i(x_i) + \frac{1}{T} \log C'_\theta(\mathbf{x})\end{aligned} \tag{14}$$

The representation is consist with two-phrase CCA framework in that the first term on the right of equation (14) implies mutual information of $\mathbf{x}$ and that the second term is empirical estimation of entropy of $\mathbf{x}$. It is not hard to proof that

$$\min D(p \| p_c) \Leftrightarrow \max \mathcal{L}(\mathbf{W}, \theta) \tag{15}$$

5.3 Estimation of Copula

Suppose the IC subspaces have been correctly determined by ICA and then we can identify the copula by minimizing the second term on the right of (7). Given a class of Copula $C(\mathbf{u}; \theta)$ with parameter vector $\theta \in \Theta$, and a set of sources $\mathbf{s} = (s_1, \ldots, s_n)$ identified from data set $\mathbf{X}$, the problem is such a optimization one

$$\max_{\mathbf{W}, \theta} E_{p(\mathbf{s})}(C'(\mathbf{u_s}; W, \theta)) \tag{16}$$

By using *Sklar* theorem, the copula to be identified has been separate with marginal distributions which are known except non-Gaussianity in ICA model. Therefore, the problem here is a semi-parametric one and only need identifying the copula.

Parametric method is adopted. First, we should select a hypothesis for copula among many types of copula available. The selection depends on many factors, such as priori knowledge, computational ease, and individual preference. Due to space limitations, only few of them are introduced here. For more detail please refer to [9].

When a set of sources $\mathbf{s}$ and a parametric copula $C(\cdot; \theta)$ is prepared, the optimization of (16) becomes an optimization problem which can be solved as follows:

$$\sum_{s_i=1}^{n} \frac{\partial C'}{\partial \theta}(\mathbf{u}; \theta) = 0 \tag{17}$$

where many readily methods can be utilized.

6 Simulation Experiments

In this section, simulation experiments are designed to compare CCA and ICA on two typical cases to investigate whether CCA can perform better than ICA as previous stated. One cases is with independent components and the other is where there are dependent components.

We first apply both methods on independent components recovery from their mixtures and then on recovery of components with dependency structure. In both experiments, the basic case of BSS with two components are considered. Two components are generated by bi-variate distribution associated with Gumbel copula:

$$C(u,v) = \exp\left(\left((-\ln u)^\theta + (-\ln v)^\theta \right)^{-\theta} \right) \tag{18}$$

where $\theta = 1, 5$ respectively. Note that two components such generated are independent when $\theta = 1$ and thus compose of sources of ICA problem. The marginal density of components are uniform. Sources are mixed by randomly generated and invertible 2×2 matrix $\mathbf{A}$. In our experiments, $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} 0.4936 \; 0.9803 \\ 0.4126 \; 0.5470 \end{pmatrix}$$

Both ICA and CCA are used to recover the components from their mixtures. Without the attention to study model selection, Gumbel copula family is adopted in CCA method.

The results are illustrated in Figure 1. Due to space limitations, we only present copula density structure of sources and their recoveries by both methods

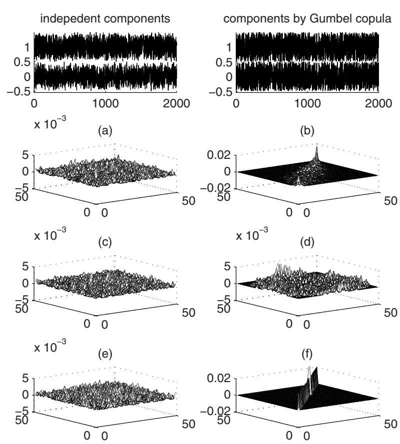

Fig. 1. Simulation experiments. The left column is for independent component experiments and the right column is for the experiment of components by Gumbel copula. The top two sub-figure is sources and (a) and (b) is their corresponding copula density. (c) and (d) is for ICA and (e) and (f) is for CCA.

in Figure 1. Note that copula density structure should be a plane if two components are independent, that is, $C(u, v) = 1$. It can be learned from figure 1 that both methods works well when components are mutually independent and more importantly that ICA always try to extracts components mutually independent while CCA can recover the dependency between components successfully.

7 Conclusions and Further Directions

In this paper, a framework named Copula Component Analysis for blind source separation is proposed as a generalization of ICA. It differs from ICA which assumes independence of sources that the underlying components may be dependent with certain structure which is represented by Copula. By incorporating dependency structure, much accurate estimation can be made, especially in the case where the assumption of independence is invalidated. A two phrase inference method is introduced for CCA which is based on the notion of multidimensional ICA. A preliminary simulated experiment demonstrates the advantage of CCA over ICA on dependency structure discovery. Many problems remain to be studied in the future, such as Identifiability of the method, selection of copula model and applications.

References

1. Comon, P.: Independent component analysis - A new concept? Signal Processing 36, 287–314 (1994)
2. Bell, A., Sejnowski, T.: An information-maximization approach to blind separation and blind deconvolution. Neural Comp. 7, 1129–1159 (1995)
3. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind source separation. In: Advances in Neural Information Processing, pp. 757–763. MIT Press, Cambridge (1996)
4. Cardoso, J.-F., Laheld, B.H.: Equivariant adaptive source separation. IEEE Transactions on Signal Processing 44, 3017–3030 (1996)
5. Pham, D.T., Garat, P.: Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. IEEE Transactions on Signal Processing 45, 1712–1725 (1997)
6. Cardoso, J.-F.: Multidimensional independent component analysis, Acoustics, Speech, and Signal Processing. In: ICASSP'98. Proceedings of the 1998 IEEE International Conference on, vol. 4, pp. 1941–1944 (1998)
7. Bach, F.R., Jordan, M.I.: Beyond independent components: Trees and clusters. Journal of Machine Learning Research 4, 1205–1233 (2003)
8. Hyvärinen, A., Hoyer, P.O., Inki, M.: Topographic Independent Component Analysis. Neural Computation 13, 1527–1558 (2001)
9. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics. Springer, New York (1999)